Sukkur IBA Journal of Computing and Mathematical Sciences

Recognized by HEC Pakistan in "Y" Category

E-ISSN: 2522-3003

P-ISSN: 2520-0755

Volume: 5 | No: 1 | Jan - Jun | 2021

Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS) is the bi-annual research journal published by **Sukkur IBA University**, Sukkur Pakistan. **SJCMS** is dedicated to serve as a key resource to provide practical information for the researchers associated with computing and mathematical sciences at global scale.

Copyright: All rights reserved. No part of this publication may be produced, translated or stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying and/or otherwise the prior permission of publication authorities.

Disclaimer: The opinions expressed in **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)** are those of the authors and contributors, and do not necessarily reflect those of the journal management, advisory board, the editorial board, Sukkur IBA University press or the organization to which the authors are affiliated. Papers published in **SJCMS** are processed through double blind peer-review by subject specialists and language experts. Neither the **Sukkur IBA University** nor the editors of **SJCMS** can be held responsible for errors or any consequences arising from the use of information contained in this journal, instead errors should be reported directly to corresponding authors of articles.

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Mission Statement

The mission of **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)** is to provide a premier interdisciplinary platform to researchers, scientists and practitioners from the field of computing and mathematical sciences for dissemination of their findings and to contribute in the knowledge domain.

Aims & Objectives

Sukkur IBA Journal of Computing and Mathematical Sciences aims to publish cutting edge research in the field of computing and mathematical sciences.

The objectives of **SJCMS** are:

1. to provide a platform for researchers for dissemination of new knowledge.
2. to connect researchers at global scale.
3. to fill the gap between academician and industrial research community.

Research Themes

The research focused on but not limited to following core thematic areas:

Computing:

- Software Engineering
- Formal Methods
- Human Computer Interaction
- Information Privacy and Security
- Computer Networks
- High Speed Networks
- Data Communication
- Mobile Computing
- Wireless Multimedia Systems
- Social Networks
- Data Science
- Big data Analysis
- Contextual Social Network Analysis and Mining
- Crowdsource Management
- Ubiquitous Computing
- Distributed Computing
- Cloud Computing
- Intelligent devices
- Security, Privacy and Trust in Computing and Communication
- Wearable Computing Technologies
- Soft Computing
- Genetic Algorithms
- Robotics
- Evolutionary Computing
- Machine Learning

Mathematics:

- Applied Mathematical Analysis
- Mathematical Finance
- Applied Algebra
- Stochastic Processes

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Volume 5 | No. 1 | January– June 2021

Patron's Message

Sukkur IBA University has been imparting education with its core values merit, quality, and excellence since its inception. Sukkur IBA University has achieved numerous milestones in a very short span of time that hardly any other institution has achieved in the history of Pakistan. The university is continuously being ranked as one of the best university in Pakistan by Higher Education Commission (HEC). The distinct service of Sukkur IBA University is to serve the rural areas of Sindh and also underprivileged areas of other provinces of Pakistan. Sukkur IBA University is committed to serve targeted youth of Pakistan who is suffering from poverty and deprived of equal opportunity to seek quality education. Sukkur IBA University is successfully undertaking its mission and objectives that lead Pakistan towards socio-economic prosperity.

In continuation of endeavors to touch new horizons in the field of computing and mathematical sciences, Sukkur IBA University publishes an international referred journal. Sukkur IBA University believes that research is an integral part of modern learnings and development. Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS) is the modest effort to contribute and promote the research environment within the university and Pakistan as a whole. SJCMS is a peer-reviewed and multidisciplinary research journal to publish findings and results of the latest and innovative research in the fields, but not limited to Computing and Mathematical Sciences. Following the tradition of Sukkur IBA University, SJCMS is also aimed at achieving international recognition and high impact research publication in the near future.

Prof. Mir Mohammad Shah
Vice Chancellor, Sukkur IBA University
Patron SJCMS

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Editorial

Dear Readers,

It is a pleasure to present to you the Sixth issue of (volume 3, issue 2) of Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS).

The stunning advances in various fields of science and technology have a profound impact on our lives in almost every sphere of our activity, such as health, agriculture, communication, transportation, and defense. These advances have been driven by an ever-growing volume of exciting discoveries, largely emanating from research community. In order to highlight the future technology challenges, the SJCMS aims to publish cutting-edge research in the field of computing and mathematical sciences for dissemination to the largest stakeholders. SJCMS has achieved milestones in very short span of time and is indexed in renowned databases such as DOAJ, Google Scholar, DRJI, BASE, ROAD, CrossRef and many others. SJCMS is now HEC recognized in Z-Category.

This issue contains the double-blind peer-reviewed articles that address the key research problems in the specified domain The SJCMS adopts all standards that are a prerequisite for publishing high-quality research work. The Editorial Board and the Reviewers Board of the Journal is comprised of renowned researchers from technologically advanced countries. The Journal has adopted the Open Access Policy without charging any publication fees that will certainly increase the readership by providing free access to a wider audience.

On behalf of the SJCMS, I welcome the submissions for upcoming issue (Volume-5, Issue-2, July-December 2021) and looking forward to receiving your valuable feedback.

Sincerely,

Dr. Ahmed Waqas
Chief Editor

Volume 5 | No. 1 | January– June 2021

Patron ***Prof. Mir Muhammad Shah***
Chief Editor ***Dr. Ahmed Waqas***
Editor ***Dr. Abdul Rehman Gilal***
Associate Editors ***Dr. M. Abdul Rehman, Dr. Javed Hussain Brohi***
Managing Editors ***Prof. Dr. Pervez Memon, Dr. Sher Muhammad Daudpota***

Editorial Board

Prof. Dr. Abdul Majeed Siddiqui
Pennsylvania State University, USA

Prof. Dr. Gul Agha
University of Illinois, USA

Prof. Dr. Muhammad RidzaWahiddin
International Islamic University, Malaysia

Prof. Dr. Tahar Kechadi
University College Dublin, Ireland

Prof. Dr. Paolo Bottoni
Sapienza - University of Rome, Italy

Prof. Dr. Md. Anwar Hossain
University of Dhaka, Bangladesh

Dr. Umer Altaf
KAUST, Kingdom of Saudi Arabia

Prof. Dr. Farid Nait Abdesalam
Paris Descartes University Paris, France

Prof. Dr. Asadullah Shah
International Islamic University, Malaysia

Prof. Dr. Adnan Nadeem
Islamia University Madina, KSA

Dr. Jafreezal Jaafar
Universiti Teknologi PETRONAS

Dr. Zulkefli Muhammad Yusof
International Islamic University, Malaysia

Dr. Hafiz Abid Mahmood
AMA International University, Bahrain

Prof. Dr. Luiz Fernando Capretz
Western University Canada

Prof. Dr. Samir Iqbal
University of Texas Rio Grande Valley, USA

Prof. Dr. S.M Aqil Burney
IoBM, Karachi, Pakistan

Prof. Dr. Zubair Shaikh
Muhammad Ali Jinnah University, Pakistan

Prof. Dr. Mohammad Shabir
Quaid-i-Azam University Islamabad, Pakistan

Dr. Ferhana Ahmad
LUMS, Lahore, Pakistan

Dr. Asghar Qadir
Quaid-e-Azam University, Islamabad

Dr. Nadeem Mahmood
University of Karachi, Pakistan

Engr. Zahid Hussain Khand
Sukkur IBA University, Pakistan

Dr. Qamar Uddin Khand
Sukkur IBA University, Pakistan

Dr. Syed Hyder Ali Muttaqi Shah
Sukkur IBA University, Pakistan

Dr. Muhammad Ajmal Sawand
Sukkur IBA University, Pakistan

Dr. Niaz Hussain Ghumro
Sukkur IBA University, Pakistan

Dr. Zarqa Bano
Sukkur IBA University, Pakistan

Dr. Javed Ahmed Shahani
Sukkur IBA University, Pakistan

Dr. Ghulam Mujtaba Shaikh
Sukkur IBA University, Pakistan

Prof. Dr. Florin POPENTIU VLADICESCU
University Politehnica in Bucharest

Language Editor - **Prof. Ghulam Hussain Manganhar**
Project and Production Management
Ms. Suman Najam Shaikh, Ms. Rakhi Batra, Mr. Hassan Abbas, Mr.Muhammad Asim Bhutto

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Vol. 5, No. 1 | January – June 2021

Breast Cancer Detection via Global and Local Features using Digital Histology Images

Ghulam Murtaza[1], Ainuddin Wahid Abdul Wahab[1], Ghulam Raza[2], Liyana Shuib[1]

Abstract:

Globally, breast cancer (BC) is the prevailing cause of unusual deaths in women. Breast tumor (BT) is a primary symptom and may lead to BC. Digital histology (DH) image modality is a gold standard medical test for a definite diagnosis of BC. Traditionally, DH images are visually examined by two or more pathologists to come up with a consensus for authentic BC detection which may cause a high error rate. Therefore, researchers had developed automated BC detection models using a machine learning (ML) based approach. Thus, this study aims to develop a BC detection model through ten feature extraction methods which extract both local and global type features from publicly available breast histology dataset. The extracted features are sorted by their weights, which are computed by the neighborhood component analysis method. A feature selection algorithm is developed to find the minimum number of discriminating features, evaluated through seven heterogeneous traditional ML classifiers. The proposed ML-based BC detection model acquired 90% accuracy for the initial testing set using 51 Harris features. Whereas, for the extended testing set, only three Harris features is shown 93% accuracy. The proposed BC detection model can assist the doctor in giving a second opinion.

Keywords: *breast cancer detection, histology image, cancer classification, machine learning, deep learning.*

1. Introduction

Breast cancer is one of the most frequent causes of death in women than in other types of cancer. According to the American Cancer Society (2018) report, around 316120 cases were diagnosed with BC, and 40610 women died unusually. Usually, BC is initiated when breast cells propagate abnormally i.e., out of control. Consequently, such abnormal cell growth forms a lump, also known as a tumor. Breast tumors (BT) are of two types, non-carcinoma and carcinoma. Moreover, non-carcinoma BT has two basic subtypes, normal and benign. Benign tumors are categorized as non-cancerous and usually do not lead to any severe health problems. Whereas in situ and invasive are subtypes of carcinoma BT, causes BC and create severe health issues. In situ is non-invasive and does not spread to the other organs of a woman's body, and remains in the mammary ductal-lobular system. In contrast, invasive BT tends to invade the surrounding breast tissues. However, if BC is timely diagnosed and pursued with proper treatment, it can be cured by 85% at the initial stage and decreased by up to 10% for later stages (Wang, Khosla, Gargeya, Irshad, & Beck,

[1] Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia
[2] Our Lady of Lourdes hospital Drogheda Ireland.
Corresponding Authors: gmurtaza@siswa.um.edu.my, ainuddin@um.edu.my

2016). In routine checkups, BC can be detected by screening examination even before any physical symptoms like a lump. Initially, detection and diagnosis of BC can be made via radiology image modalities such as mammograms (MGs), magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and thermography. However, MGs are commonly adopted for the last three decades for BC screening examination (Tabar et al., 2011). In the case of abnormal tissue growth found in breast screening examination, biopsy techniques are recommended to ensure the presence of BC. The most popular biopsy techniques are fine needle aspiration biopsy, core needle biopsy, vacuum-assisted biopsy, incisional biopsy, and excisional biopsy(BREASTCANCER.ORG, 2015). In a biopsy procedure, the sample of cells or tissues is collected from the affected area of the breast and placed over a glass microscope slide for dying and further microscopic analysis. The visual microscopic study of the histology samples allows both cell and tissue level examination of the breast, which enables a pathologist to detect BC presence assertively. Thus, histology samples are used as the gold standard for BC detection as well as for classification (Rubin, Strayer, & Rubin, 2008). Recently, due to the advent of digital pathology labs, the microscopic histology slides are scanned and transformed into digital images called whole slide images (WSI). The cell and tissue level examination in WSI enabled pathologists to mark regions of interest (ROI) (i.e., cancerous and non-cancerous areas). The ROI is marked based on interpretations made for overall tissues' architectures by examining cell distribution, nuclei organization, density, shape (i.e., regular or irregular), and variations in stained tissues(Gurcan et al., 2009). For a detailed analysis of each ROI, various image patches are captured from marked ROI of WSI using a camera with various zooming factors (i.e., 40x, 100x, 200x and 400x)(Spanhol, Oliveira, Petitjean, & Heutte, 2016b) named as digital histology/histopathological (DH) images. The final decision of BC detection and classification is made with the consensus of interpretations made by two or more expert pathologists after adopting rigorous facial/visual examination of DH images. However, the manual facial analysis of DH images is an inefficient time task. Moreover, the interpretation accuracy of DH image diagnosis depends upon the domain knowledge (Bige et al., 2011), training, experience, and differences in circumstances and personal interest of learning (Desjardins, 1960) of a pathologist. However, a minor misinterpretation among BC types may lead to significant health and financial loss. Therefore, there is an urge for a machine learning (ML) based automated computerized system to assist the pathologist for better interpretation of DH images of BC.

The ML detection from BC histology tissues is characterized by two types of features, namely local and global features. The global features are also called structural features, include contour, edge density, direction, nuclei shape, and glandular unit shape representations like histogram-oriented gradients (HOG) features. Whereas local features (also named texture features) represent the spatial distribution of intensity (rough, smooth, silky, or bumpy) in a region of an image like local binary patterns (LBP). Harris and Stephens (1988) proposed a combined corner and edge detection algorithm based on the local auto-correlation function. The auto-correlation function had eliminated the limitation of a shifting window by a small amount of variation proposed by Moravec (1980). The correlation function not only detects the corners and edges but also measures the edge quality by selecting isolated corner pixels for thinning the edge pixels. It is achieved by taking three measures; first, all small window shifts are made through analytic expansion about the shift origin. Second, a smooth circular window was adopted like Gaussian instead of a binary and rectangular window. Third, the corner measure can use both edges with direction-shift instead of depending on minimum edge value. Thus, the Harris method based features are more accurate than the Moravec method due to less number of false positives to detect edges and corners. Jianbo and Tomasi (1994) developed the eigenvalue algorithm to detect the corners of an object. Unlike Harris features using the response function score, the author used minimum eigenvalue (MinEigen) to decide corners. It can be mathematically expressed by $T(v) = \lambda v$, where λ eigenvalue associated with eigenvector v. Geometrically eigenvalue is a point stretched by a transformation in a direction by some non-zero factor called eigenvector. Ojala, Pietikäinen, and Mäenpää (2002) presented the LBP method to recognize certain binary patterns named as uniform. The uniform patterns are basic characteristics of local image texture, and their histogram is an authenticated source of dominant texture features. The proposed method computes a generalized gray-scale and rotation invariant operator to detect uniform patterns. The extracted uniform patterns are resilient to any quantization of the angular space and for any spatial resolution. Moreover, it combines multiple operators for multi-resolution analysis. Thus, the developed approach is highly robust to grayscale variation and rotation. Moreover, this approach is computationally simple, efficient because operators can be realized with few comparisons in a small neighbourhood and a lookup table. Such capabilities make LBP method more applicable to time critical applications. Matas, Chum, Urban, and Pajdla (2004) proposed an efficient detection algorithm for an affinely invariant stable subset of external regions termed as maximally stable extremal regions (MSER). The external regions are set of image elements when put into correspondence possess two vital properties; for instance, these sets are closed under the continuous transformation of image coordinates and monotonic transformation of image intensities. MSER algorithm extracts features from brighter or darker regions distinct from their surroundings. Whereas, these regions are stable by their relevant range of thresholds of the intensity function. Therefore, MSER can extract regions more efficiently, like crosswalk regions or regions with distinct illumination variations. Dalal and Triggs (2005) presented work on HOG features to predict pedestrian detection in still images. The basic idea is the distribution of intensity gradients or edge directions, which can identify the appearance and shape of an object in an image. Therefore, HOG features are also categorized as global or structural features. It is achieved by splitting an image into cells i.e., small overlapping regions. Afterward, a histogram for gradient direction of each cell is computed to form a single HOG feature. Ultimately, a combination of all cells HOG features represents the object present in the entire image. Moreover, the intensity of an image block (group of cells) is calculated for histogram contrast normalization of each cell present in respective block. This normalization enhances the HOG features accuracy and robustness to alterations in illumination and shadowing. Rosten and Drummond (2005) introduced real-time tracking system by collective use of both points and edges properties present in an image. For real time feature tracking the author proposed features from accelerated segment test (FAST) corner feature detector technique. FAST techniques use 16 pixels to detect target pixel p belongs to a corner pixel. Thus, each pixel which surrounds p in a circle is labelled in clockwise fashion by values starting from 1 to 16, see Fig 1. If a set of continuous surrounded pixels N in a circle are brighter than the intensity of p plus threshold value t or all darker than the intensity of p minus threshold value t, then p is detected as corner pixel. Therefore, FAST technique will show good results if three or more N are available around p. Otherwise, there is a need to add blur with normal distribution to get better corner detection using FAST technique. Bay, Ess, Tuytelaars, and Van Gool (2008) coined rotation invariant detectors and features named as speed-up robust feature (SURF). These can be computed faster and outperformed (by scale variant feature transform - SIFT) due to distinctiveness and robustness. This is achieved by using integral images for image convolutions, Hessian matrix approximation for the detection, sums of Haar wavelet components for feature extraction, along with simplifying these methods as needed. SURF is a faster method due to the use of integral images and Laplacian indexing for the matching step. Some of the effective applications are object recognition, image classification, reconstruction, content-based image retrieval, and registration. Leutenegger, Chli, and Siegwart (2011) proposed a binary robust invariant scalable keypoint (BRISK) method for keypoint detection, description, and matching. This method had proved adaptive, better quality algorithm on benchmark datasets by showing low power consumptions compared to SURF. The success lies in the use of unique scale-space derived from FAST detection algorithm along with building a bit-string descriptor by comparing the intensities of the keypoint neighbourhood. Thus, BRISK uses an easily computed circular sampling pattern for brightness comparison to create a binary descriptor string. Due to its unique properties, it is useful for many real-time applications where limited time with low power consumption matters like a content-based image revival. Rublee, Rabaud, Konolige, and Bradski (2011) proposed oriented FAST and Rotated BRIEF(ORB)as a cost-effective and faster than SURF extraction method based on BRIEF. Moreover, this method is also rotation invariant and robust to noise. To achieve this, the first keypoint extracted by FAST. Afterwards, the top point is selected by Harris corner detection. Moreover, an efficient orientation component is added to the FAST feature algorithm. This makes ORB as rotation invariant compared to FAST.KAZE features are named to tribute the father of scale-space analysis (Weickert, Ishikawa, & Imiya, 1999). Alcantarilla, Bartoli, and Davison (2012) proposed the KAZE feature extraction method, which involves various steps. For instance, initially, nonlinear scale space is computed using an efficient Additive Operator Splitting (AOS) technique with variable conductance diffusion. Next, using nonlinear scale-space the response of scale-normalized determinant of Hessian at multiple levels is computed to detect 2D features of interest. Finally, rotation-invariant features were extracted by approximating the dominant orientation found in local neighbourhood centred at the keypoint location for first order image derivatives.

The main contributions of this study are given below.

1. This study develops an efficient (consumes less training and detection time), cost-effective (can be trained on a normal desktop computer), and feasible (requires a smaller number of labelled images to show good results) model for BC detection using DH images.

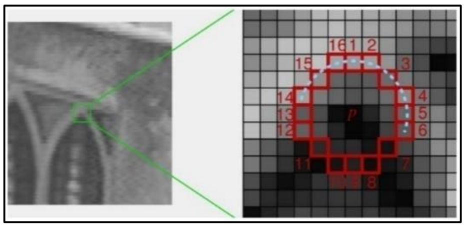


Fig 1: FAST feature extraction technique(Rosten & Drummond, 2005)

2. Ten (local and global) feature extraction methods (i.e., HOG, MinEigen, Harris, LBP, KAZE, MSER, SURF, FAST, BRISK, and ORB) are evaluated to get the most distinct features.

3. A feature selection algorithm is developed to extract the minimum number of discriminative feature subsets to complement classification results.

4. The selected reduced number of features are evaluated via six performance evaluation metrics, namely accuracy, sensitivity, specificity, precision, F-measure, and Matthews correlation coefficient using seven heterogeneous traditional ML classifiers, namely SVM, NB, LR, LDA, kNN, GK, and DT.

5. The results of the proposed BC detection model (complemented with feature reduction algorithm) are compared with ten state-of-the-art baseline models.

Further sections of this study are arranged as follows. Section 2 gives an overview of current literature, limitations of existing BC detection and classification models, and the main contributions of this study. The overall research methodology employed is discussed in section 3. Whereas Section 4 discusses the experimental setups and subsequent results. Next, section 5 discussion covers the significance of experimental results, limitations of this research with future directions. Finally, section 6 conclusion recaps the overall research made in this study.

2. Literature Review

In the past three decades, most of the research had been focused on radiology images (Tabar et al., 2011) especially mammogram (breast x-rays) images using traditional machine learning (ML) based approaches for binary classification (Hassanien & Ali, 2006; Kontos & Maragoudakis, 2013; Biswas, Nath, & Roy, 2016; Nusantara, Purwanti, & Soelistiono, 2016; Ponraj, Poongodi, & Mercy, 2017). For instance, Ponraj et al. (2017) performed BC detection using MG images of a publicly available dataset. The author extracted textural features using LBP and local gradient pattern (LGP) along with related histograms. The extracted feature set is classified through SVM, where LGP outperformed the LBP by showing better accuracy, sensitivity, and specificity as 95%, 92%, and 94%, respectively. Moreover, Nusantara et al. (2016) used digital MGs to classify them into normal and abnormal categories. Where ROI from MGs was cropped to extract features like energy, mean, and standard deviation from the third level of wavelet decomposition coefficients. The highest reported accuracy, sensitivity, and specificity were 96.8%, 100%, and 95%, respectively. Similarly, Biswas et al. (2016) utilized the MGs of a public dataset for normal and abnormal tissue classification. The author extracted texture features from image ROIs using a gray level co-occurrence matrix (GLCM) followed by many pre-processing steps like removal of artifacts, noise reduction, and contrast enhancement. The author achieved a higher 95% accuracy, 100% sensitivity, and 90% specificity through SVM compared to kNN and artificial neural networks. Apart from a binary classification using radiology images, many researchers used DH images to solve the BT multiclassification problem via deep learning (DL) based approaches (Araujo et al., 2017; Han et al., 2017; Wan, Cao, Chen, & Qin, 2017; Zheng et al., 2017; Murtaza, Shuib, Mujtaba, & Raza, 2019; Murtaza, Shuib, Wahab, et al., 2019; Murtaza, Shuib, Abdul Wahab, et al., 2020; Murtaza, Shuib, Wahab, et al., 2020). For instance, Han et al. (2017) proposed a class structure-based deep convolutional neural network (CSDCNN) which embeds the non-linear representation learning model instead of handcrafted features. CSDCNN was initially trained on imageNet dataset to show better results for the BreakHis dataset. The author reported accuracy for eight subtypes of BT was 93.2%. Moreover, Zheng et al. (2017) proposed a nucleus guided feature extraction method. Initially, the proposed deep convolution network (DCNN) was trained using nucleus features without labels; afterward, the trained DCNN was fined tuned and retrained using labelled DH images of an exclusive dataset. The reported accuracy for the 15-classes dataset was 96.4%. A summary of proposed methodology and limitations of aforementioned studies given in Table I.

It has been observed from the aforementioned studies that the traditional ML-based BC detection models are better than DL based models due to five reasons. First, DL based BC detection models require high computational resources like GPU, RAM with longer training time compared to traditional ML models. Second, DL models need a large number of images to avoid overfitting issue. Whereas, traditional ML model can be trained efficiently using a small number of images to show comparable results. Third, it is highly difficult to optimize the hyper-parameters of deep layered networks before initiating the training process. Fourth, DL based models get very smaller size input images like AlexNet(Krizhevsky, Sutskever, & Hinton, 2012) to get 227x227 pixels, whereas DH images usually are of very high resolution like BCBH DH image size is 2040 x 1536 pixels. Thus, rescaling is mandatory before feeding into a deep neural network, which causes loss of information (Komura & Ishikawa, 2018). Hence, DL based models are more suitable to solve the multiclassification problem where a larger number of labelled images are available with ample computational resources. However, the traditional ML-based BC detection model is a better and worthwhile choice for BC detection using DH images.

In this study, DH images are utilized for BC detection confidently compared to radiology images. Moreover, ten aforementioned feature extraction methods (FEMs) namely HOG, MinEigen, Harris, LBP, KAZE, MSER, SURF, FAST, BRISK, and ORB are used to extract local and global features from DH images. Both, local and global types of features are helping to choose the most distinct features using various ML classifiers to detect BC accurately from DH images. Moreover, many studies used a similar type of FEMs, thus provides direct results comparison of the proposed classification approach with state-of-the-art baseline models. Next, the extracted features from each FEMs are stored in the master feature table (MFT). Where, a row in MFT represents many features of each image present in the training set, and two (initial and extended) testing sets. Thus, ten MFTs are created to represent each of the FEM. Finally, MFTs are classified through seven heterogeneous traditional ML learning classifiers namely support vector machine (SVM), naïve Bayes (NB), linear regression (LR), linear discriminant analysis (LDA), k nearest neighbor (kNN), Gaussian Kernel (GK) and decision tree (DT) using six performance evaluation metrics such as accuracy, sensitivity, specificity, precision, F-measure and Matthews correlation coefficient.

Table I: A summary of the methodology and their limitations of existing models for breast cancer detection

Reference	Methodology	Limitations
Nusantara et al. (2016)	ROI-based wavelet decomposition is used for feature extraction and classification is made by kNN.	The result can be compromised if ROI is not extracted properly. Thus, have a higher dependency of mammogram expert.
Biswas et al. (2016)	Gray Level Co-occurrence Matrix (GLCM) used on ROI of a mammogram of MIAS dataset. kNN(k=3) is used for classification.	The result can be compromised if ROI is not extracted properly. Thus, have a higher dependency of mammogram expert.
Ponraj et al. (2017)	Features extracted via local binary pattern (LBP) and local gradient pattern (LGP) and classified through SVM.	Uses a small dataset thus model training can be biased. Needs to show other PEMs like F-measure or MCC to show class level biasness.
Han et al. (2017)	Proposed CNN-based model for classification of BreakHis dataset for 8 subtypes of BrT using histopathology images.	The model requires a large number of images for pretraining like ImageNet. Thus, needs GPU and longer training time.
Zheng et al. (2017)	A CNN-based model trained on histopathology images where the nucleus of cancer lesion is already marked.	Needs a histopathology image expert to mark the nucleus of the cancer lesion.

3. Proposed Method

This section demonstrates the employed overall research methodology to develop the BC classification model for carcinoma and non-carcinoma types. Four stages (see Fig 2)are involved namely data collection, classification models training, and classification with evaluation. The tasks performed in each phase are discussed in detail in the subsequent sections.

3.1. Data Collection

In the first stage of the research methodology (see Fig 2) a publicly available bioimaging challenge 2015 breast histology (BCBH) dataset (Teresa Araújo, 2015) is utilized to develop the BC cancer detection model. This dataset is hosted by the Institute for Systems and Computer Engineering, Technology, and Science (INESC TEC) associated laboratory. BCBH dataset possesses breast histology, annotated, uncompressed images with 2040 x 1536 pixels' resolution. Whereas, all the images were acquired with the magnification of 200x with the same protocol. Moreover, the dataset contains H&E stained images divided into four types of BT, namely normal, benign, in situ, and invasive. Normal and benign subjects are assumed as non-cancerous (i.e., non-carcinoma), whereas the rest of the two subtypes (i.e., in situ and invasive) are cancerous (i.e., carcinoma). Thus, in this study BC detection is made as carcinoma and non-carcinoma basic types of a tumor. BCBH corpus possesses overall 285 DH images (see Table 2). Where 249 out of 285 images are provided for model training purposes, and the rest of the images (i.e., 36) are designated for model testing. Moreover, the testing set is divided into two parts, named as an initial testing set and an extended testing set. The initial testing set possesses 20, whereas the extended testing set consists of 16 DH images. The images in both training and testing sets are almost equally distributed for each class label. BCBH dataset is publicly accessible at URL[3] .

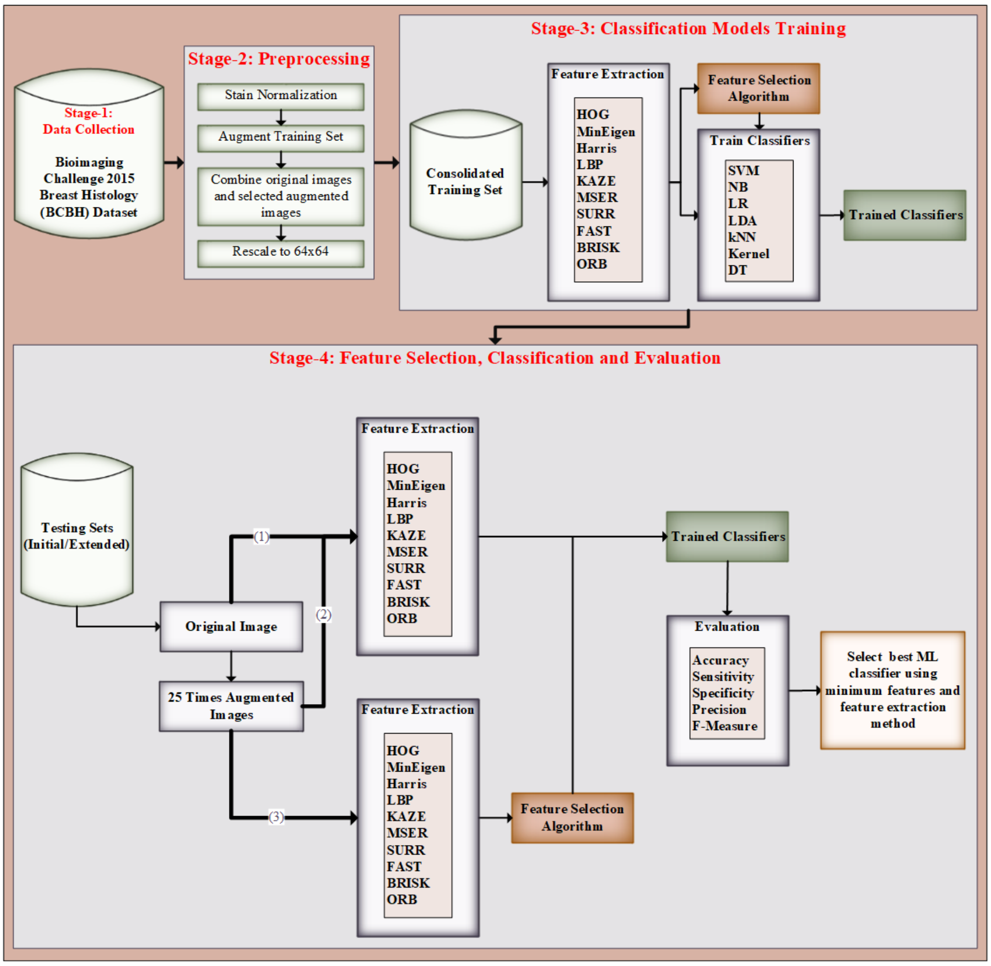


Fig 2: Overall research methodology. Feature selection strategies (1) Original image, (2) Augmented images, and (3) Augmented images with feature selection algorithm.

3 https://rdm.inesctec.pt/dataset/nis-2017-003.

Table II. Distribution of images in bioimaging challenge 2015 breast histology dataset

		Non-carcinoma	Carcinoma	Total
		Normal + Benign	In situ + Invasive	
Training Set		55+69	63+62	249
Testing Sets	Initial	5+5	5+5	20
	Extended	4+4	4+4	16
Total		132	143	285

3.2. Image pre-processing

In the second stage of research methodology (see Fig 2), the image pre-processing consists of many necessary tasks required to reduces the misclassification rate by cleaning the raw data before initiating classifier training as well as testing processes. In this study, the adopted pre-processing tasks are stain normalization, augmentation, and the rescaling of images.

3.2.1. Stain normalization

DH images inherently possess higher inconsistencies due to two major reasons. First, the cumbersome process of making biopsy slides for microscopic facial analysis. Second, the scanning process of biopsy slides into digital images to create WSI. The preparation of biopsy slides involves pathology lab protocols, human skills, embedding, sectioning, and coloring (McCann, Ozolek, Castro, Parvin, & Kovacevic, 2015). Moreover, the preparation of WSI in digital pathological labs involves various scanners (to convert biopsy microscopic slides into digital WSI) of many vendors, which possess different scanning protocols. Thus, inconsistencies (i.e., visible variability) are found due to different resolution, color, brightness, and contrast settings of scanners. Therefore, to avoid such dominant variabilities which may largely distract the overall classification process, need to be eliminated before initiating the training and testing process for DH images. In this study, Reinhard, Adhikhmin, Gooch, and Shirley (2001) method is adopted to harmonize the stain of DH image. However, there are many stain normalization methods (Macenko et al., 2009; Khan, Rajpoot, Treanor, & Magee, 2014), but these methods are weaker to preserves the BC lesion structures. Thus, the use of such methods may cause a loss of important information to detect carcinoma lesions.

3.2.2. Image augmentation

The image augmentation is performed to enhance the performance of traditional ML classifiers for the collected BCBH dataset. In this study, the training set has been augmented by twenty-five times using basic image processing techniques like rotation, flip, transform, and padding. An algorithm (see Fig 3) is developed to generate twenty-five synthetic images from a single original image by applying the aforementioned basic image processing techniques. Each image of the training set is augmented by twenty-five times and overall, 6225 (i.e., 249x25) augmented training images are generated. However, out of 6225, 3100 belong to non-carcinoma, whereas 3125 are carcinoma augmented images. Whereas, the original images of the training set are also used along with augmented images for training purposes. The overall training set possesses 3224 (i.e., 3100+55+69) non-carcinoma and 3250 (i.e., 3125+63+62) carcinoma images, which are utilized for training of ML classifiers. Moreover, the developed augmentation algorithm is also used for testing sets to get further analysis of each augmented image to classify the original testing set image.

3.2.3. Image rescaling

BCBH dataset DH images are larger like 2040 x 1536 pixels. Thus, it needs lots of computational time to extract the features. In this study, it has been experimentally examined that the larger image size does not make a major effect on the extraction of meaningful features. Besides, the translation of images in the image augmentation algorithm returns an arbitrary size. Thus, rescaling is required to make all images of the same size before applying any FEM. This pre-processing step makes the entire classification process resources efficient without affecting the overall performance of the BC classification.

3.3. Classification models training

In stage three of the research methodology (see Fig 2), two tasks are carried out using the augmented training set. First, feature extraction using various feature extraction techniques. Second, training of multiple ML classifiers using two approaches. In the first approach, the training of ML classifiers is performed by using all extracted features. Whereas, the second approach reduces the number of features by implementing the proposed feature selection algorithm. The details of each task are given in the subsequent sections.

3.3.1. Feature Extraction

The preprocessed training set is utilized to extract features by ten FEMs, namely HOG, MinEigen, Harris, LBP, KAZE, MSER, SURF, FAST, BRISK, and ORB. Many studies have used a similar type of FEMs; thus, it provides an opportunity for direct comparison of results for the proposed classification model. The extracted features of each FEM method are stored in MFT. Where a row in MFT represents many features of each image of training and two testing sets. Thus, ten MFTs are created to represent each of the aforementioned FEM.

3.3.2. ML classifiers training

In this study, seven heterogeneous traditional ML classifiers, namely SVM, NB, LR, LDA, kNN, GK, and DT are trained through aforesaid ten MFTs. Because most of the classifiers are used by baseline studies, that allows direct comparison of results of the proposed BC detection model. Here, the ML classifiers' training task is based on two approaches. First, to train above mentioned seven classifiers using all MFTs. Second, the use of the proposed feature selection algorithm (see **Error! Reference source not found.**) to reduce the features for the training of the above-given ML classifiers.

```
Input: OrgImg
Output: AugImg(25)
Function ImageAugmenter
    define array AugImg(25)

    AugImg(1) = flip(Img,TopToBottom)
    AugImg(2) = flip(Img,LeftToRight)

    AugImg(3) = rotate(AugImg(1),90)
    AugImg(4) = rotate(AugImg(2),90)

    AugImg(5) = translate(AugImg(1),H/2,-W/2)
    AugImg(6) = translate(AugImg(1),H/2, W/2)
    AugImg(7) = translate(AugImg(1),-H/2,-W/2)
    AugImg(8) = translate(AugImg(1),-H/2, W/2)

    AugImg(9) = translate(AugImg(2),H/2,-W/2)
    AugImg(10) = translate(AugImg(2),H/2, W/2)
    AugImg(11) = translate(AugImg(2),-H/2,-W/2)
    AugImg(12) = translate(AugImg(2),-H/2, W/2)

    AugImg(13) = transform(OrgImg, affine([1 0 0; 1 1 0; 0 0 1]))
    AugImg(14) = transform(OrgImg, affine([1 0.33 0; 0 1 0; 0 0 1]))
    AugImg(15) = transform(OrgImg, affine([1 0.9 0; 0 1 0; 0 0 1]))
    AugImg(16) = transform(OrgImg affine([1 0 0; .5 1 0; 0 0 1]))

    AugImg(17) = transform(AugImg(1), affine([1 0 0; 1 1 0; 0 0 1]))
    AugImg(18) = transform(AugImg(1), affine([1 0.33 0; 0 1 0; 0 0 1]))
    AugImg(19) = transform(AugImg(1), affine([1 0.9 0; 0 1 0; 0 0 1]))
    AugImg(20) = transform(AugImg(1), affine([1 0 0; .5 1 0; 0 0 1]))

    AugImg(21) = transform(AugImg(2), affine([1 0 0; 1 1 0; 0 0 1]))
    AugImg(22) = transform(AugImg(2), affine([1 0.33 0; 0 1 0; 0 0 1]))
    AugImg(23) = transform(AugImg(2), affine([1 0.9 0; 0 1 0; 0 0 1]))
    AugImg(24) = transform(AugImg(2), affine([1 0 0; .5 1 0; 0 0 1]))

    AugImg(25) = Padding(OrgImg,affine[1 0; 1 1; 0 0]/2])
end
```

Fig 3: Image augmentation algorithm

The proposed feature selection algorithm performs two tasks in a cascade manner. First, it sorts (in descending order) the features of MFT of each feature extraction method by computing feature weights through neighborhood component analysis (NCA), see Fig 5, shows the sample graphs of HOG features only, whereas the same approach is implemented for all FEMs to get rid of those features which have a very low contribution (i.e., weight) to classify carcinoma and non-carcinoma distinctly. Thus, NCA weights are computed for the MFT of each feature extraction method.Fig 5 first column, shows the sample graphs of HOG unsorted features. Next, all MFTs' columns are sorted in descending order by NCA weights, (see Fig 5 second column) the sample graphs of HOG sorted features. The sorted MFT is referred to hereafter as SMFT. Finally, the top 64 features from all SMFTs are used for further analysis. Because it has been experimentally observed, the classification results are at their maximum level while using the top 59

```
Input: TrainingSet,TestingSetInitial,TestingSetExtended
Output: BestClassifier
Function FeatureSelection
        Agument TrainingSet,TestingSetInitial,TestingSetExtended by ImageAugmenter
        Extract Features TrainingSet,TestingSetInitial,TestingSetExtended using ten methods
        Create MFT_Training,MFT_Testing_Initial,MFT_Testing_Extended for ten feature extraction methods
        Comments:  compute weights by neighbourhood component analysis (NCA)
        FeatureWeights = nca(MFT_Training,Labels_Training)
        Ftr_Idx  = sort(FeatureWeights,'desc')
        Comments: training set features by NCA weight
        SMFT_Training = MFT_Training(:,Ftr_Idx)
        Comments: initial testing set features by NCA weight
        SMFT_Testing_Initial = MFT_Testing_Initial(:,Ftr_Idx)
        Comments:  sort extended testing set features by NCA weight
        SMFT_Testing_Extended = MFT_Testing_Extended(:,Ftr_Idx)
        Comments:  classify using reduced features by 1,3,5,7,.....
        for Ftr = 1:TotalNoFeatures step by 2
                Train SVM, NB, LR, LDA, kNN, GK and DT using SMFT_Training(:,Ftr)
                Classify SVM, NB, LR, LDA, kNN, GK and DT using SMFT_Testing_Initial(:,Ftr)
                Classify SVM, NB, LR, LDA, kNN, GK and DT using SMFT_Testing_extended(:,Ftr)
                Evaluate SVM, NB, LR, LDA, kNN, GK and DT using accuracy, sensitivity and F-measure
        end
        Select best feature extraction method, best classifier with minimum feature subset
end
```

Fig 4: Feature selection algorithm

features of each SMFT. Fig 5 third column, shows the sample graphs of HOG 64 features of the highest NCA weights. Moreover, the proposed feature selection algorithm takes features subset (i.e., 1,3,5, 7…) incrementally from SMFT to train the aforementioned seven ML classifiers. The goal of the proposed feature selection algorithm is to get a minimum feature subset to train the classifier without compromising the overall classification performance. In summary, Fig 5 first column graphs display the NCA weights (unsorted) and the second column represents the NCA weights (sorted in descending order) of overall HOG features. However, the third column graphs of Fig 5 show the selection of the minimum number of features based on higher weights. It has been experimentally observed that the classification results are at their maximum level while using the top 59 out of 1764 NCA weighted HOG features.

3.4. Feature selection, classification, and evaluation

In the fourth stage of the overall research methodology (see Fig 2), the classification and evaluation of two testing sets (i.e., initial and extended) are made by following three strategies of feature selection. In the first feature selection strategy, the original images of both testing sets are used to extract features using the aforementioned ten FEMs. Here, all features are used for classification and evaluation purposes. Whereas, in the second feature selection strategy, the features are extracted from ten methods followed by applying twenty-five augmentation methods individually on all original images of both testing sets. For classification and evaluation, the overall extracted features of each image augmentation method are used for results analysis. However, in the third strategy of feature selection, the proposed feature selection algorithm is implemented, followed by all steps mentioned in the second strategy. In contrast, to both first and second feature selection strategies, the classification and evaluation are performed using selected reduced features subset (i.e., 1,3,5, 7…). After adopting the three feature selection strategies, the selected features are used for classification using aforesaid seven trained ML classifiers. Whereas, the classification results are evaluated through six performance evaluation metrics (PEMs) like accuracy, sensitivity, specificity, precision, F-measure, and Matthews correlation coefficient. Finally, the top-performing ML classifier using a minimum number of features is selected for each of the two testing sets.

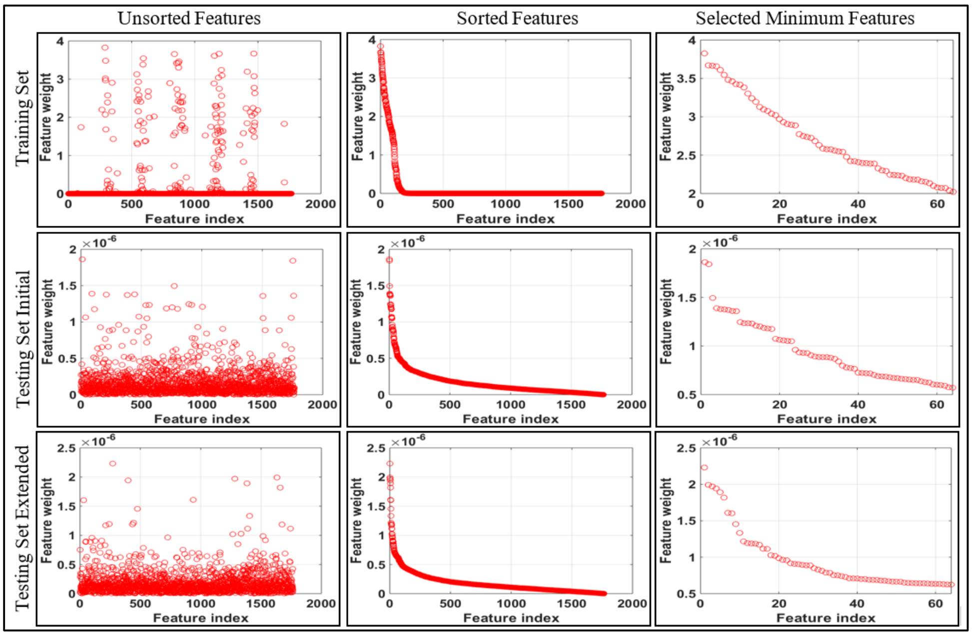


Fig 5. NCA weights of extracted HOG features for the training set and two testing sets

4. Experimental setup and analysis

The experiments are organized to assess the performance of ten FEMs namely HOG, MinEigen, Harris, LBP, KAZE, MSER, SURF, FAST, BRISK, and ORB through seven traditional ML classifiers namely SVM, NB, LR, LDA, kNN, GK, and DT. This research involves three experimental settings, see Fig 6

I. **Experiments using original images:** In the first experimental setting, the performance of ten FEMs through seven ML classifiers is analysed using two testing sets namely initial and extended. Finally, all the results are reported in this study.

II. **Experiments using augmented images:** In the second experimental setting, each original image is augmented twenty-five times. Afterward, each augmented image is used to evaluate the performance of ten FEMs through seven ML classifiers using two testing sets. Finally, the best results among twenty-five augmentation methods are shown in this study.

III. **Experiments using augmented images with feature selection algorithm:** In the third experimental setting, thirty-two feature subset (1,3, 5….63) with twenty-five augmentation methods are evaluated by showing the performance of ten FEMs through seven ML classifiers using two testing sets. Finally, the best results with reduced features are shown for further analysis.

All classification model development steps like data splitting, image pre-processing, ML classifiers training, testing, and evaluation are performed in MATLAB R2019a. Furthermore, all experiments are carried out using default parameters except those which are specifically mentioned in this study.

4.1. Classification Performance evaluation metrics

In this study, the overall experimental results analysis is made through three main PEMs namely accuracy (Ac), sensitivity (Sn), F-measure (Fm), and Matthews Correlation Coefficient (MCC) see Fig 7. Because Ac (Fig 7, equation 1) is the most

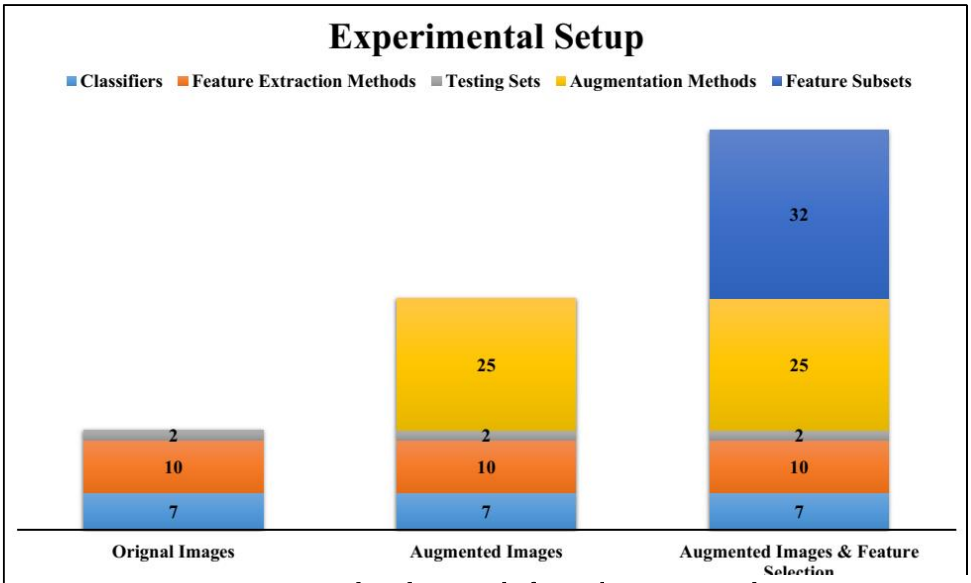
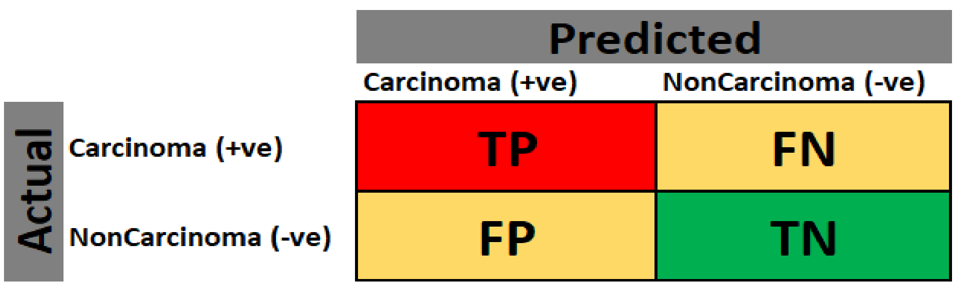


Fig 6. Experimental analysis made for each experimental setup



$$Ac = (TP+TN)/(TP+FN+FP+TN) \quad (1)$$

$$Sn = TP/(TP+FN) \quad (2)$$

$$Sp = FP/(FP+TN) \quad (3)$$

$$Pr = TP/(TP+FP) \quad (4)$$

$$F\text{-}Measure = (2*Pr*Sn)/(Pr+Sn) \quad (5)$$

$$MCC = (TP*TN-FP*FN)/[(TP+FP)*(FN+TN)*(FP+TN)*(TP+FN)]^{\wedge}(1/2) \quad (6)$$

Fig 7. PEMs (like Ac, Sn, Sp, Pr, F-Measure and MCC) equations adopted for experimental results analysis

common PEMs needed to compare baseline studies, while Sn (Fig 7, equation 2) is highly important in medical science diagnosis of cancer (i.e., cancer positive cases). However, the other PEMs like specificity (Sp) (Fig 7, equation 3), and precision (Pr) (Fig 7, equation 4) are also reported along with four outcomes of the confusion matrix (see Fig 7) namely true positive (TP), false negative (FN), false positive (FP), and true negative (TN) in the appendix-A (see Table 1, Table 2, and Table 3). Whereas, Fm (Fig 7, equation 5) and MCC (Fig 7, equation 6) can evaluate the quality (i.e., biasedness) of classifier predictions. MCC involves the overall outcomes of the confusion matrix. It returns a coefficient value between +1 and -1.

Where, coefficient value +1, 0, and -1 represents the perfect, average, and reverse predictions of a classifier respectively. Thus, in this study, MCC is adopted to select the top-performing classifier with unbiased predictions.

4.2. Experimental Results Analysis

In this section, the results of above mentioned three experimental setups are reported for analysis. The BC classification results are discussed in terms of four (i.e., Ac, Sn, Fm, and MCC) PEMs.

4.2.1. Experimental results using original images

This section reports the results of seven (i.e., SVM, NB, LR, LDA, kNN, GK, and DT) ML classifiers using ten (i.e., HOG, MinEigen, Harris, LBP, KAZE, MSER, SURF, FAST, BRISK, and ORB) FEMs for two (i.e., Initial and Extended) testing sets. In the first experimental setup, overall, 140 (7 classifiers x10 FEMs x 2 testing sets) analyses are reported using four PEMs like Ac, Sn Fm, and MCC see Table 3. In Table 3 , the experimental results are split into two parts based on initial and extended testing sets. While using the initial testing set it can be observed from

Table 3 that the kNN (k=1,3,5) classifier using ORB features has outperformed by showing the highest 75.00% Ac, 78.26% Fm, and 0.5241 MCC with the second-highest Sn of 90.00%. Whereas the 100.00% Sn is achieved by kNN using HOG features with lower Ac, Sn, and MCC as 55.00%,68.00%, and 0.2294.

Furthermore, the highest accuracy of 75.00% is also achieved through kNN using LBP features but with compromised MMC as 0.5025. In contrast, NB+SURF has shown the lowest performance (i.e., Ac= 35.00% and Sn=00.00%) using the initial testing set. On the other hand, while using the extended testing set, it can be seen in

Table that the better Ac and MCC like 81.25% and 0.6742 are gained by LDA+MSER but with a lower Sn of 62.50%. Whereas, the better Sn (i.e., 100.00%) is attended by both SVM+HOG and kNN+FAST. Moreover, the top Fm value of 87.50% is shown by LR for BRISK features with compromised Ac, Sn, and MCC. Conversely, LR produces the worst results while using both SURF and HOG features i.e., Ac=50.00%, Sn=00.00%. Similarly, NB+HOG has shown poor results (i.e., Ac=50.00%, Sn=00.00%) using the extended testing set. Summarizing, ORB and MSER features have outperformed among all FEMs using initial and extended testing sets by acquiring better MCC value for BC detection.

Table III. Experimental results using original images

		Initial Testing Set							Extended Testing Set						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Ac**	**75.00**	60.00	50.00	70.00	55.00	45.00	60.00	68.75	50.00	50.00	62.50	56.25	62.50	43.75
	Sn	80.00	100.00	**0.00**	80.00	**100.00**	50.00	50.00	75.00	87.50	**0.00**	62.50	**100.00**	62.50	50.00
	Fm	76.19	71.43	NaN	72.73	68.97	47.62	55.56	70.59	63.64	NaN	62.50	69.57	62.50	47.06
	MCC	0.5025	0.3333	NaN	0.4082	0.2294	-0.1005	0.2041	0.3780	0.0000	NaN	0.2500	0.2582	0.2500	-0.1260
MinEigen	**Ac**	60.00	55.00	40.00	40.00	50.00	70.00	**35.00**	**31.25**	43.75	50.00	56.25	43.75	37.50	68.75
	Sn	40.00	20.00	30.00	20.00	50.00	70.00	50.00	25.00	25.00	25.00	50.00	50.00	12.50	62.50
	Fm	50.00	30.77	33.33	25.00	50.00	70.00	43.48	26.67	30.77	33.33	53.33	47.06	16.67	66.67
	MCC	0.2182	0.1400	-0.2041	-0.2182	0.0000	0.4000	-0.3145	-0.3780	-0.1348	0.0000	0.1260	-0.1260	-0.2887	0.3780
Harris	**Ac**	40.00	60.00	60.00	60.00	55.00	55.00	45.00	62.50	50.00	**31.25**	37.50	43.75	62.50	43.75
	Sn	40.00	30.00	40.00	40.00	50.00	70.00	50.00	75.00	50.00	25.00	37.50	37.50	50.00	25.00
	Fm	40.00	42.86	50.00	50.00	52.63	60.87	47.62	66.67	50.00	26.67	37.50	40.00	57.14	30.77

		Initial Testing Set							**Extended Testing Set**						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	MCC	-0.2000	0.2500	0.2182	0.2182	0.1005	0.1048	-0.1005	0.2582	0.0000	-0.3780	-0.2500	-0.1260	0.2582	-0.1348
LBP	**Ac**	65.00	60.00	70.00	60.00	**75.00**	60.00	65.00	62.50	68.75	62.50	68.75	62.50	68.75	37.50
	Sn	80.00	70.00	90.00	80.00	70.00	80.00	80.00	50.00	50.00	50.00	75.00	62.50	50.00	25.00
	Fm	69.57	63.64	75.00	66.67	73.68	66.67	69.57	57.14	61.54	57.14	70.59	62.50	61.54	28.57
	MCC	0.3145	0.2041	0.4364	0.2182	0.5025	0.2182	0.3145	0.2582	0.4045	0.2582	0.3780	0.2500	0.4045	-0.2582
KAZE	**Ac**	55.00	55.00	55.00	50.00	65.00	45.00	50.00	56.25	56.25	56.25	43.75	43.75	56.25	43.75
	Sn	20.00	20.00	20.00	50.00	50.00	20.00	50.00	12.50	25.00	12.50	25.00	37.50	12.50	50.00
	Fm	30.77	30.77	30.77	50.00	58.82	26.67	50.00	22.22	36.36	22.22	30.77	40.00	22.22	47.06
	MCC	0.1400	0.1400	0.1400	0.0000	0.3145	-0.1155	0.0000	0.2582	0.1601	0.2582	-0.1348	-0.1260	0.2582	-0.1260
MSER	**Ac**	55.00	70.00	65.00	60.00	55.00	70.00	60.00	62.50	68.75	62.50	**81.25**	75.00	56.25	56.25
	Sn	20.00	90.00	40.00	30.00	30.00	40.00	40.00	37.50	62.50	37.50	**62.50**	75.00	25.00	25.00
	Fm	30.77	75.00	53.33	42.86	40.00	57.14	50.00	50.00	66.67	50.00	**76.92**	75.00	36.36	36.36
	MCC	0.1400	0.4364	0.3464	0.2500	0.1155	0.5000	0.2182	0.2887	0.3780	0.2887	**0.6742**	0.5000	0.1601	0.1601
SURF	**Ac**	60.00	**35.00**	40.00	45.00	55.00	65.00	50.00	56.25	56.25	50.00	43.75	56.25	56.25	56.25
	Sn	80.00	**0.00**	**0.00**	**0.00**	10.00	80.00	10.00	50.00	12.50	**0.00**	12.50	37.50	50.00	37.50
	Fm	66.67	NaN	NaN	NaN	18.18	69.57	16.67	53.33	22.22	NaN	18.18	46.15	53.33	46.15
	MCC	0.218218	-0.42008	-0.3333	-0.22942	0.229416	0.314485	0.0000	0.125988	0.258199	NaN	-0.16013	0.13484	0.125988	0.13484
FAST	**Ac**	70.00	60.00	60.00	60.00	50.00	45.00	50.00	62.50	62.50	50.00	50.00	62.50	56.25	43.75
	Sn	90.00	20.00	20.00	20.00	10.00	70.00	20.00	**100.00**	62.50	25.00	37.50	62.50	62.50	25.00
	Fm	75.00	33.33	33.33	33.33	16.67	56.00	28.57	72.73	62.50	33.33	42.86	62.50	58.82	30.77
	MCC	0.4364	0.3333	0.3333	0.3333	0.0000	-0.1155	0.0000	0.3780	0.2500	0.0000	0.0000	0.2500	0.1260	-0.1348
BRISK	**Ac**	65.00	50.00	50.00	50.00	55.00	60.00	45.00	50.00	62.50	56.25	68.75	62.50	43.75	68.75
	Sn	80.00	10.00	20.00	20.00	20.00	60.00	10.00	87.50	75.00	25.00	62.50	62.50	50.00	62.50
	Fm	69.57	16.67	28.57	28.57	30.77	60.00	**15.39**	**12.50**	50.00	**87.50**	75.00	62.50	37.50	75.00
	MCC	0.3145	0.0000	0.0000	0.0000	0.1400	0.2000	-0.1400	0.0000	0.2582	0.1601	0.3780	0.2500	-0.1260	0.3780
ORB	**Ac**	50.00	50.00	55.00	50.00	**75.00**	55.00	40.00	62.50	43.75	50.00	50.00	50.00	37.50	50.00
	Sn	50.00	60.00	40.00	40.00	90.00	50.00	50.00	75.00	37.50	50.00	50.00	50.00	62.50	37.50
	Fm	50.00	54.55	47.06	44.44	**78.26**	52.63	45.46	66.67	**40.00**	50.00	50.00	50.00	50.00	42.86
	MCC	0.0000	0.0000	0.1048	0.0000	0.5241	0.1005	-0.2041	0.2582	-0.1260	0.0000	0.0000	0.0000	-0.2887	0.0000

4.2.2. Experimental results analysis using image augmentation algorithm

This section represents the results of the aforementioned seven ML classifiers for ten FEMs using twenty-five image augmentation methods for two (i.e., Initial and Extended) testing sets. In the second experimental setup, overall, 140 (7 classifiers x10 FEMs x 2 testing sets x 1 Augmentation method) analyses of top-performing augmentation method is reported using three PEMs like Ac, Sn, and Fm, see Table 4.

Table , by analysing the results of the initial testing set, it can be observed that the highest 0.7035 MCC with comparable Sn (i.e.; 90.00%) is shown by DT+MSER. However, similar MCC is also achieved by NB+MSER but with a compromised 80.00%. Whereas, the highest Sn (i.e., 100.00%) is produced by SVM in most of the FEMs with lower MCC. Contrary, NB, and LR have shown the lowest performance (i.e., Ac=50.00%, Sn=00.00%) using BRISK features. On the flip side, using extended testing set the better Ac (i.e., 90.00%), Sn (i.e., 100.00%), and MCC (i.e., 0.8165) are achieved by SVM+BRISK with a lower Fm value of 80.00%. Whereas, a better Fm of 100.00% is observed using both LDA and DT with BRISK features. In contrast, the lowest Ac (i.e., 50.00%) and Sn (i.e., 00.00%) are noticed with both SURF and FAST features when classified through LR. In summary, the MSER and BRISH feature outperformed among all FEMs using initial and extended testing sets by acquiring better MCC values for BC detection.

Table IV. Experimental results analysis using augmented images

		Initial Testing Set							**Extended Testing Set**						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Ac**	65.00	75.00	70.00	80.00	75.00	70.00	75.00	81.25	62.50	56.25	75.00	75.00	75.00	43.75
	Sn	90.00	90.00	**100.00**	70.00	90.00	90.00	80.00	87.50	**100.00**	37.50	87.50	75.00	62.50	50.00
	Fm	72.00	78.26	76.92	77.78	78.26	75.00	76.19	82.35	72.73	46.15	77.78	75.00	71.43	47.06
	MCC	0.3464	0.5241	0.5000	0.6124	0.5241	0.4364	0.5025	0.6299	0.3780	0.1348	0.5164	0.5000	0.5164	-0.1260
MinEigen	**Ac**	65.00	70.00	80.00	80.00	85.00	65.00	70.00	50.00	62.50	50.00	87.50	81.25	75.00	62.50
	Sn	**100.00**	**100.00**	70.00	90.00	70.00	70.00	80.00	**100.00**	75.00	87.50	87.50	75.00	87.50	50.00
	Fm	74.07	76.92	77.78	81.82	82.35	66.67	72.73	66.67	66.67	63.64	87.50	80.00	77.78	57.14
	MCC	0.4201	0.5000	0.6124	0.6124	0.7338	0.3015	0.4082	NaN	0.2582	0.0000	0.7500	0.6299	0.5164	0.2582
Harris	**Ac**	75.00	60.00	65.00	65.00	65.00	75.00	70.00	68.75	81.25	50.00	75.00	81.25	62.50	37.50
	Sn	**100.00**	90.00	80.00	80.00	90.00	90.00	70.00	**100.00**	**100.00**	62.50	62.50	87.50	87.50	37.50
	Fm	80.00	69.23	69.57	69.57	72.00	78.26	70.00	76.19	84.21	55.56	71.43	82.35	70.00	37.50
	MCC	0.5774	0.2500	0.3145	0.3145	0.3464	0.5241	0.4000	0.4804	0.6742	0.0000	0.5164	0.6299	0.2887	-0.2500
LBP	**Ac**	65.00	70.00	70.00	70.00	80.00	70.00	75.00	68.75	62.50	43.75	81.25	81.25	68.75	62.50
	Sn	**100.00**	90.00	90.00	90.00	90.00	90.00	**100.00**	**100.00**	87.50	50.00	87.50	100.00	**100.00**	75.00
	Fm	74.07	75.00	75.00	75.00	81.82	75.00	80.00	76.19	70.00	47.06	82.35	84.21	76.19	66.67
	MCC	0.4201	0.4364	0.4364	0.4364	0.6124	0.4364	0.5774	0.4804	0.2887	-0.1260	0.6299	0.6742	0.4804	0.2582
KAZE	**Ac**	70.00	75.00	80.00	85.00	80.00	80.00	75.00	75.00	68.75	50.00	75.00	68.75	68.75	50.00
	Sn	**100.00**	80.00	80.00	**100.00**	70.00	70.00	80.00	75.00	75.00	**100.00**	75.00	100.00	75.00	**100.00**
	Fm	76.92	76.19	80.00	86.96	77.78	77.78	76.19	75.00	70.59	66.67	75.00	76.19	70.59	66.67
	MCC	0.5000	0.5025	0.6000	0.7338	0.6124	0.6124	0.5025	0.5000	0.3780	NaN	0.5000	0.4804	0.3780	NaN
MSER	**Ac**	65.00	85.00	70.00	75.00	70.00	80.00	**85.00**	68.75	68.75	50.00	81.25	87.50	75.00	**18.75**
	Sn	80.00	80.00	70.00	70.00	70.00	70.00	90.00	62.50	87.50	37.50	62.50	87.50	87.50	25.00
	Fm	69.57	84.21	70.00	73.68	70.00	77.78	**85.71**	66.67	73.68	42.86	76.92	87.50	77.78	**23.53**
	MCC	0.3145	0.7035	0.4000	0.5025	0.4000	0.6124	0.7035	0.3780	0.4045	0.0000	0.6742	0.7500	0.5164	-0.6299
SURF	**Ac**	75.00	75.00	75.00	75.00	70.00	80.00	70.00	56.25	62.50	50.00	68.75	75.00	62.50	50.00
	Sn	90.00	90.00	80.00	90.00	80.00	90.00	60.00	87.50	87.50	**0.00**	75.00	87.50	87.50	**0.00**
	Fm	78.26	78.26	76.19	78.26	72.73	81.82	66.67	66.67	70.00	NaN	70.59	77.78	70.00	NaN
	MCC	0.5241	0.5241	0.5025	0.5241	0.4082	0.6124	0.4082	0.1601	0.2887	NaN	0.3780	0.5164	0.2887	NaN
FA	**Ac**	80.00	65.00	60.00	60.00	65.00	75.00	70.00	75.00	81.25	50.00	68.75	75.00	68.75	50.00

		Initial Testing Set							Extended Testing Set						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	Sn	**100.00**	30.00	20.00	20.00	40.00	80.00	40.00	**100.00**	87.50	**0.00**	50.00	75.00	87.50	**0.00**
	Fm	83.33	46.15	**33.33**	**33.33**	53.33	76.19	57.14	80.00	82.35	NaN	61.54	75.00	73.68	NaN
	MCC	0.6547	0.4201	0.3333	0.3333	0.3464	0.5025	0.5000	0.5774	0.6299	NaN	0.4045	0.5000	0.4045	NaN
BRISK	**Ac**	75.00	75.00	50.00	81.25	75.00	68.75	50.00	**90.00**	50.00	60.00	65.00	55.00	80.00	65.00
	Sn	**100.00**	75.00	**0.00**	75.00	62.50	87.50	**0.00**	**100.00**	40.00	30.00	30.00	40.00	**100.00**	30.00
	Fm	80.00	75.00	NaN	80.00	71.43	73.68	NaN	**80.00**	60.00	90.00	**100.00**	70.00	60.00	**100.00**
	MCC	0.5774	0.5000	NaN	0.6299	0.5164	0.4045	NaN	**0.8165**	0.0000	0.2500	0.4201	0.1048	0.6547	0.4201
ORB	**Ac**	85.00	65.00	60.00	60.00	70.00	70.00	65.00	62.50	68.75	43.75	81.25	68.75	68.75	43.75
	Sn	80.00	70.00	80.00	80.00	90.00	80.00	80.00	62.50	75.00	37.50	75.00	75.00	75.00	50.00
	Fm	84.21	66.67	66.67	66.67	75.00	72.73	69.57	62.50	70.59	40.00	80.00	70.59	70.59	47.06
	MCC	0.7035	0.3015	0.2182	0.2182	0.4364	0.4082	0.3145	0.2500	0.3780	-0.1260	0.6299	0.3780	0.3780	-0.1260

4.2.3. Experimental results using image augmentation algorithm with feature selection algorithm

This section demonstrates the results of the aforementioned seven traditional ML classifiers using ten FEMs and twenty-five image augmentation methods followed by minimum features selection for two (i.e., Initial and Extended) testing sets. In the third experimental setup, overall, 140 (7 classifiers x10 FEMs x 2 testing sets x 1 Augmentation method x 1 reduced features subset) analyses of top-performing classifiers using reduced features are reported using four PEMs like Ac, Sn, Fm, and MCC.

Initial testing set results in Table shows that GK outperformed using the least number of Harris features (i.e., 517 out of 64) with the highest Ac (i.e., 90%), Sn (i.e., 90%), Fm (i.e., 95.24%) and MCC (i.e., 0.80). However, the same results are produced via SVM by using 57 out of 1764 HOG features. Besides, HOG and LBP features have shown 100% Sn using most of the classifiers. Other classifiers like NB (i.e., 3 out of 64 FAST features) and LDA (i.e., 19 out of 64 FAST features) have shown the worst performance by achieving the lowest Sn and Fm like 30% and 46.15%. Contrastingly, for extended testing, set the results in Table illustrate that kNN got better results (i.e., Ac=93.75, Sn=100%, Fm=94.12%, and MCC=0.8819) using 3 out of 1764 Harris features. Moreover, 100% Sn is observed using HOG, Harris, and ORB reduced features using almost all classifiers. Moreover, approximately the same results are obtained via kNN+MinEigen by using 32 out of 64 features. In contrast, LR using (13 out 64) BRISK features has shown the lowest Sn (i.e., 37.50%) and compromised Ac (i.e., 68.75%) using the extended testing set. In a nutshell, Harris and FAST selective features have outperformed among all FEMs using initial and extended testing set by acquiring better MCC using the least number of features for BC detectio

Table V. Experimental results using augmented images with the feature selection algorithm

		Initial Testing Set							Extended Testing Set						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Ac**	**90.00**	75.00	85.00	85.00	85.00	80.00	85.00	75.00	81.25	81.25	81.25	93.75	87.50	87.50
	Sn	**90.00**	**100.00**	**100.00**	**100.00**	**100.00**	90.00	80.00	**100.00**	**100.00**	**100.00**	87.50	87.50	**100.00**	87.50
	Fm	**90.00**	80.00	86.96	86.96	86.96	81.82	84.21	80.00	84.21	84.21	82.35	93.33	88.89	87.50

		Initial Testing Set							Extended Testing Set						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	MCC	**0.8000**	0.5774	0.7338	0.7338	0.7338	0.6124	0.7035	0.5774	0.6742	0.6742	0.6299	0.8819	0.7746	0.7500
	Features	**57**	3	61	63	5	11	29	23	1	61	1	27	3	1
MinEigen	**Ac**	80.00	75.00	80.00	80.00	90.00	85.00	80.00	81.25	75.00	87.50	87.50	**93.75**	93.75	87.50
	Sn	**100.00**	**100.00**	70.00	90.00	80.00	80.00	80.00	**100.00**	87.50	**100.00**	75.00	**87.5**	87.50	87.50
	Fm	83.33	80.00	77.78	81.82	88.89	84.21	80.00	84.21	77.78	88.89	85.71	**93.33**	93.33	87.50
	MCC	0.6547	0.5774	0.6124	0.6124	0.8165	0.7035	0.6000	0.6742	0.5164	0.7746	0.7746	**0.8819**	0.8819	0.7500
	Features	49	29	7	63	11	13	39	19	13	17	13	**23**	19	55
Harris	**Ac**	85.00	70.00	80.00	75.00	80.00	**90.00**	80.00	87.50	81.25	81.25	81.25	**93.75**	87.50	87.50
	Sn	**100.00**	**100.00**	90.00	90.00	**100.00**	**90.00**	90.00	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	87.50
	Fm	86.96	76.92	81.82	78.26	83.33	**90.00**	81.82	88.89	84.21	84.21	84.21	**94.12**	88.89	87.50
	MCC	0.7338	0.5000	0.6124	0.5241	0.6547	**0.8000**	0.6124	0.7746	0.6742	0.6742	0.6742	**0.8819**	0.7746	0.7500
	Features	55	7	41	19	29	51	3	39	3	3	3	**3**	9	4
LBP	**Ac**	75.00	70.00	80.00	75.00	85.00	75.00	80.00	75.00	75.00	75.00	87.50	81.25	81.25	87.50
	Sn	**100.00**	**100.00**	**100.00**	90.00	80.00	**100.00**	**100.00**	**100.00**	87.50	87.50	87.50	**100.00**	**100.00**	75.00
	Fm	80.00	76.92	83.33	78.26	84.21	80.00	83.33	80.00	77.78	77.78	87.50	84.21	84.21	85.71
	MCC	0.5774	0.5000	0.6547	0.5241	0.7035	0.5774	0.6547	0.5774	0.5164	0.5164	0.7500	0.6742	0.6742	0.7746
	Features	3	1	5	4	7	3	10	6	9	3	3	3	6	7
KAZE	**Ac**	80.00	85.00	90.00	90.00	90.00	85.00	85.00	81.25	81.25	81.25	81.25	87.50	75.00	87.50
	Sn	**100.00**	90.00	90.00	**100.00**	90.00	**100.00**	90.00	87.50	75.00	87.50	**100.00**	**100.00**	75.00	**100.00**
	Fm	83.33	85.71	90.00	90.91	90.00	86.96	85.71	82.35	80.00	82.35	84.21	88.89	75.00	88.89
	MCC	0.6547	0.7035	0.8000	0.8165	0.8000	0.7338	0.7035	0.6299	0.6299	0.6299	0.6742	0.7746	0.5000	0.7746
	Features	3	21	7	51	11	3	61	53	13	39	43	17	11	9
MSER	**Ac**	80.00	85.00	80.00	80.00	80.00	90.00	85.00	68.75	68.75	75.00	81.25	87.50	75.00	87.50
	Sn	90.00	90.00	90.00	80.00	80.00	80.00	90.00	75.00	87.50	62.50	75.00	87.50	**100.00**	**100.00**
	Fm	81.82	85.71	81.82	80.00	80.00	88.89	85.71	70.59	73.68	71.43	80.00	87.50	80.00	88.89
	MCC	0.6124	0.7035	0.6124	0.6000	0.6000	0.8165	0.7035	0.3780	0.4045	0.5164	0.6299	0.7500	0.5774	0.7746
	Features	9	43	9	13	5	19	47	5	29	21	25	17	9	11
SURF	**Ac**	75.00	75.00	80.00	85.00	80.00	75.00	80.00	75.00	68.75	68.75	68.75	87.50	68.75	81.25
	Sn	**100.00**	**100.00**	90.00	90.00	90.00	90.00	80.00	87.50	87.50	87.50	87.50	75.00	87.50	87.50
	Fm	80.00	80.00	81.82	85.71	81.82	78.26	80.00	77.78	73.68	73.68	73.68	85.71	73.68	82.35
	MCC	0.5774	0.5774	0.6124	0.7035	0.6124	0.5241	0.6000	0.5164	0.4045	0.4045	0.4045	0.7746	0.4045	0.6299
	Features	17	49	5	17	3	5	15	3	7	3	9	11	7	41
FAST	**Ac**	85.00	60.00	65.00	65.00	70.00	80.00	70.00	75.00	87.50	75.00	81.25	81.25	87.50	81.25
	Sn	**100.00**	40.00	**30.00**	**30.00**	40.00	100.00	40.00	**100.00**	87.50	50.00	62.50	62.50	**100.00**	75.00
	Fm	86.96	50.00	**46.15**	**46.15**	57.14	83.33	57.14	80.00	87.50	66.67	76.92	76.92	88.89	80.00
	MCC	0.7338	0.2182	0.4201	0.4201	0.5000	0.6547	0.5000	0.5774	0.7500	0.5774	0.6742	0.6742	0.7746	0.6299
	Features	45	3	3	19	5	7	63	5	9	27	37	7	15	31

		Initial Testing Set							Extended Testing Set						
		SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
BRISK	**Ac**	90.00	65.00	65.00	65.00	70.00	85.00	65.00	75.00	75.00	68.75	81.25	87.50	81.25	87.50
	Sn	100.00	40.00	**30.00**	**30.00**	40.00	100.00	40.00	**100.00**	75.00	**37.50**	75.00	75.00	**100.00**	87.50
	Fm	90.91	53.33	**46.15**	**46.15**	57.14	86.96	53.33	**50.00**	75.00	**100.00**	87.50	**100.00**	62.50	87.50
	MCC	0.8165	0.3464	0.4201	0.4201	0.5000	0.7338	0.3464	0.5774	0.5000	0.4804	0.6299	0.7746	0.6742	0.7500
	Features	25	17	27	27	5	25	31	57	13	13	21	17	31	47
ORB	**Ac**	80.00	**75**	**85**	**85**	**85**	**80**	80	81.25	87.5	81.25	75	81.25	87.5	87.5
	Sn	90.00	90.00	90.00	**100.00**	90.00	80.00	80.00	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	75	75
	Fm	81.82	78.26	85.71	86.96	85.71	80.00	80.00	84.211	88.889	84.211	80	84.211	85.714	85.714
	MCC	0.6124	0.5241	0.7035	0.7338	0.7035	0.6000	0.6000	0.6742	0.7746	0.6742	0.5774	0.6742	0.7746	0.7746
	Features	1	7	11	11	31	7	19	37	33	59	9	5	29	21

In general, the BC classification results are drastically improved by proposed algorithms like image augmentation and feature selection algorithms, see Fig 8. Fig 8 (a), bar-graph represents the results for the initial testing set, whereas the Fig 8 (b) bar-graph shows the results for the extended testing set. While using the initial testing set, it can be observed from Fig 8(a) that PEMs are improved radically. Such as Ac is increased by 15% (from 75% to 90%), Fm is raised by 11.74% (from 78.26 to 90%), and MCC elevated from 0.5241 to 0.8 when augmented images are used with feature selection algorithm. Here, Sn remains stagnant like 90%, during all experiments. However, features are reduced by 20.31% i.e., only 51 out of 64 Harris features are utilized to get the highest results like 90% for each Ac, Sn, Fm. Here, the highest MCC obtained is 0.80 using the initial testing set to detect BC.

Similarly, while using an extended testing set the PEMs trend is exponentially increased, see Fig 8 (b). For instance, Ac is increased by 11.75% (from 81.25% to 93%), Sn is enhanced by 37.5% (from 62.5% to 100%), Fm value is improved by 17.2% (from 76.92% to 94.12%) when the image augmentation is applied by using feature selection algorithm. Moreover, the features are reduced by 95.31%, i.e., only 3 out of 64 Harris features are used to get maximum results like 93% Ac, 100% Sn,94.12% Fm, and 0.8819 MCC.

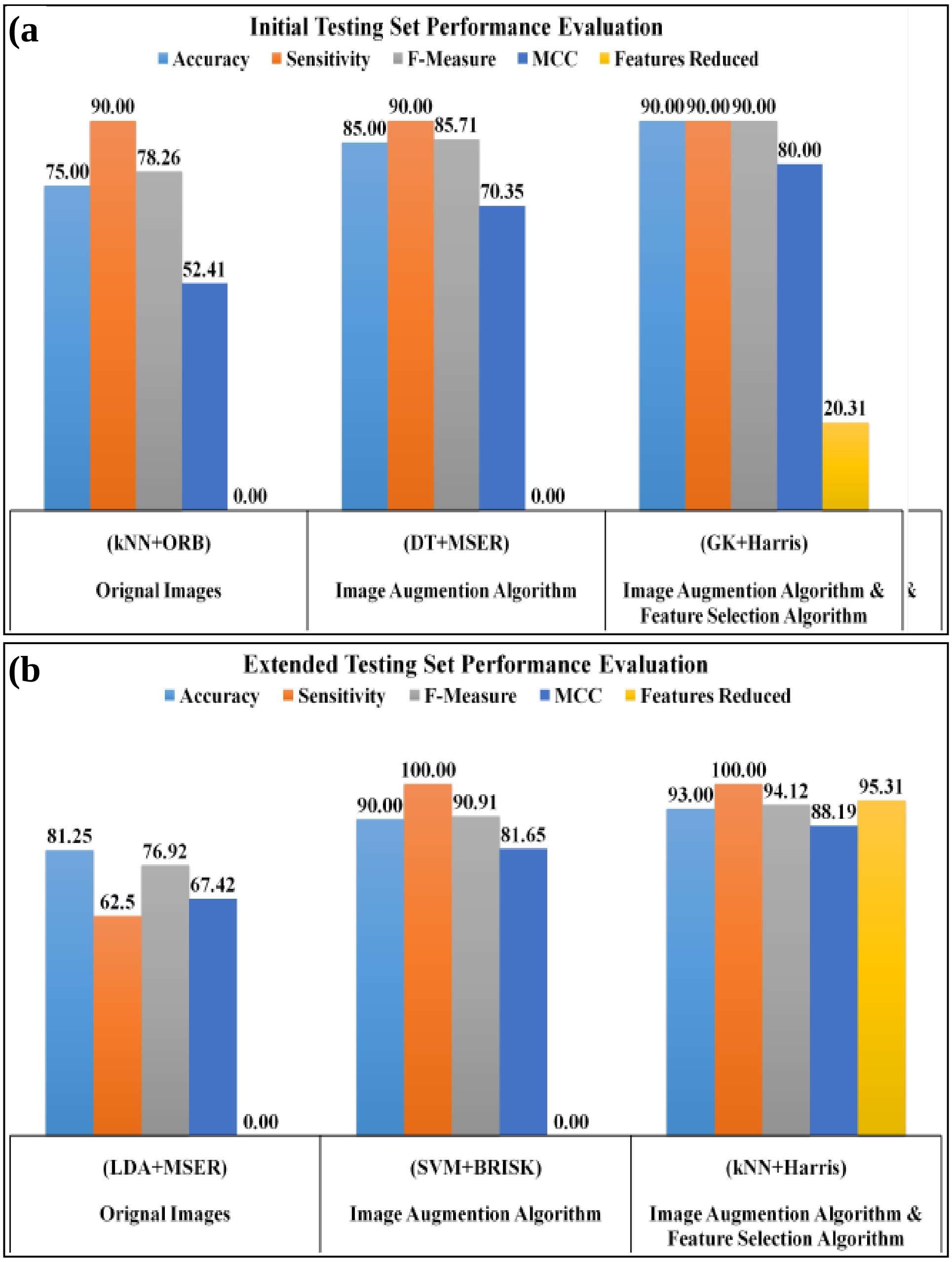


Fig 8. Three experimental setup results comparison among features of original images, augmented images and augmented images with selected features. a). Initial dataset: Three types of features ORB, MSER and Harris are used with kNN, DT and GK classifiers. b). Extended dataset: Three types of features MSER, BRISK and Harris are used with LDA, SVM and kNN classifiers.

5. Discussion

This section reveals the significance of the proposed BC detection model, developed using the traditional ML-based approach for DH images. As mentioned in the introduction section breast DH images are used for confident detection and classification of BC. For automatic detection, most of the research is made for

BC detection using DH images via DL based models and very few studies have employed traditional ML-based models. Both types of BC detection models have their pros and cons. For instance, DL based models usually require high computation power, RAM, and longer training time. Such type of deeply layered models also requires a large number of annotated images for proper training(Han et al., 2017). Moreover, it is a cumbersome task (based on the trial-and-error method) to adjust hyperparameters before initiating the training process to get the desired results. However, these models have shown good results to solve multiclassification (up to eight classes) problem (Han et al., 2017; Murtaza, Shuib, Mujtaba, et al., 2019). In contrast, the traditional ML-based approach had shown compromised results to solve the multiclassification problem. Moreover, for BC detection, the ML-based models achieved comparable results to DL based models. Apart from comparable performance, the traditional ML-based BC detection models can be trained efficiently using a normal desktop computer and do not expect a large number of labelled images to better results. Thus, it can be concluded that, apart from DL based multiclassification model, the traditional ML-based model is a feasible solution for BC detection, where we have limited computational resources and smaller datasets like medical image datasets.

Most of the existing state-of-the-art studies used DL based approach (Janowczyk & Madabhushi, 2016; Spanhol, Oliveira, Petitjean, & Heutte, 2016a; Araujo et al., 2017; Song, Zou, Chang, & Cai, 2017; Spanhol, Oliveira, Cavalin, Petitjean, & Heutte, 2017; Gupta & Bhavsar, 2018; Huang & Chung, 2018)for BC detection using DH images. For instance, Janowczyk and Madabhushi (2016) used AlexNet configuration to create a DL based model to detect invasive BC. The proposed model was trained on GPU for 22 hours using an exclusive dataset of DH images for BC detection. The reported Ac and Fm are 84.68% and 76.48%. Spanhol et al. (2016a) proposed a transferred learning-based DL model trained through the BreakHis dataset. The author acquired Ac was 85.6±4.8% using a GPU for 3 hours. Araujo et al. (2017) hosted the BCBH dataset for BT classification using DH images. The author proposed a DL based model for feature extraction and classification is made through softmax and SVM. SVM outperformed softmax by showing accuracies of 90% and 81.3% for initial and extended testing sets. However, there is room to improve the performance for BC detection mentioned in the aforementioned studies. Huang and Chung (2018) developed a CNN based spatial fused residual network using the BCBH dataset. The DH images were split into non-overlapping patches to avoid image rescaling. Moreover, a spatial relationship among patches was utilized to enhance the BC detection using high-resolution DH images. The author reported a high 98.5% accuracy and 99.6% AUC. However, the model training time was very high (i.e., 30 minutes/iteration) using GPU. Moreover, the classification duration reported was 80 minutes/image. It can be concluded from the aforementioned studies that most of the DL-based models consumed high computational resources, longer training time, and required a large number of labeled images for training, which is not a cost-effective solution for BC detection using DH images. Moreover, most of the studies only reported Ac and it can be biased towards a particular class. Thus, any other PEM like Sn, Sp, Pr or Fm is needed to be measured to show unbiased/reliable classification results. Where, Sn is highly important in medical science to identify the misclassification of true positives.

A few of the existing state-of-the-art studies (Wan, Liu, Chen, & Qin, 2014; Belsare, Mushrif, Pangarkar, & Meshram, 2015; Spanhol et al., 2016b; Gupta, Bhavsar, & Ieee, 2017) have implemented a traditional ML-based approach for BC detection using DH images. As proof, Belsare et al. (2015) used an exclusive DH image dataset for BC detection as non-malignant and malignant tissues. The author extracted GLCM, Graph Run Length Matrix (GRLM) features, and

Euler number followed by the spatio-color-texture graph segmentation method. The LDA, kNN and SVM traditional ML classifiers were employed for classification. Where LDA outperformed the others by showing an accuracy of 80% for malignancy detection. However, Belsare's BC detection model needs to be tested over any standard dataset for fare comparison of results. Spanhol et al. (2016b) hosted BreakHis, a bit larger, multifaceted, publicly available standard dataset, which carries 7909 DH images of 82 patients classified into eight subtypes of BT. The author detected BC as benign or malignant using various FEMs namely LBP, completed LBP (CLBP), local phase quantization (LPQ), GLCM, threshold adjacency statistics (TAS), and ORB. Next, the extracted features were classified via kNN, quadratic linear analysis (QDA), SVM, and random forest (RF) classifiers. Where, SVM outperformed the others by getting 85% accuracy and 86.1% AUC for binary classification. Moreover, Gupta et al. (2017) performed BC detection using the BreakHis dataset. The author created a fusion of six texture features, such as normalized color space representation, multilayer coordinate clusters representation, Gabor features on Gaussian color model, Gabor chromatic features, complex wavelet features, and chromatic features, and opponent color local binary pattern (OCLBP). Afterward, classification is made based on voting through various traditional ML classifiers, namely SVM, kNN, DT, DA, and ensemble classifiers. The reported accuracy is 88.69%.

Table , that most traditional ML-based models are unable to get better accuracy compared to DL based models. Nonetheless, DL based models consumed high computational resources, longer training time, and need a large number of annotated images to avoid overfitting issues. It is a cumbersome task to collect a large number of labelled medical images. Conversely, traditional ML-based models Table , that the proposed model achieved better Sn, and MCC, even higher than DL based models. The proposed model obtained 90% results for each Ac and Fm with MCC value 0.8 using 51 (20.31% reduced) features for the initial testing set. However, for extended testing set, the proposed model acquired 93% accuracy, 94.12% Fm, and 0.8819 MCC value using 3 (95.31% reduced) features. Further results like Pr and confusion matrix can be seen in are efficient (in terms of computational resources and time) and usually do not require a large number of labelled images, therefore suitable for medical image data. Thus, in this study, we aim to achieve better accuracy while using the least resources (normal desktop computer), training time, and a small number of labelled images. It can be seen from

Table 5 and Table 9. Thus, the proposed model is an efficient, feasible (can show better results on a small number of images), and cost-effective solution for BC automatic detection using a desktop computer. Moreover, it can be used for other applications, where time for detection maters like content-based image retrieval

.

Table VI. Performance evaluation metrics comparison of the proposed model to the state-of-the-art exiting models

Reference	**Classification approach**	**Ac (%)**	**Sn (%)**	**Class labels**	**Dataset type, name**	**Limitations**
Wan et al. (2014)	Traditional Machine Learning	87.97	86.8	Mitotic/ Non-mitotic	Public, MITOS	• Needs to improve performance.

Reference	**Classification approach**	**Ac (%)**	**Sn (%)**	**Class labels**	**Dataset type, name**	**Limitations**
Belsare et al. (2015)	Traditional Machine Learning	80	100	Non-malignant/ Malignant	Exclusive, Not given	• Needs to improve performance of Malignancy, • The use of a public dataset can show different results.
Gupta et al. (2017)	Traditional Machine Learning	88.69	---	Benign/Malignant	Public, BreakHis	• Needs to improve performance, • Sn, Fm, and MCC need to be reported
Spanhol et al. (2016b)	Traditional Machine Learning	85	---	Benign/Malignant	Public, BreakHis	• Needs to improve performance, • Sn, Fm, and MCC need to be reported.
Song et al. (2017)	Deep Learning	86.2 ± 3.7	---	Benign/Malignant	Public, BreakHis	• Needs to improve performance, • Sn, Fm, and MCC need to be reported, • It requires high computational resources and training time.
Spanhol et al. (2017)	Deep Learning	84.2 ± 1.7	---	Benign/Malignant	Public, BreakHis	• Needs to improve performance, • Sn, Fm, and MCC need to be reported, • It requires high computational resources and training time.
Gupta and Bhavsar (2018)	Deep Learning	95.9 ± 4.2	---	Benign/Malignant	Public, BreakHis	• Only Ac is reported which can be biased.

Reference	**Classification approach**	**Ac (%)**	**Sn (%)**	**Class labels**	**Dataset type, name**	**Limitations**
						• Sn, Fm, and MCC need to be reported, • Requires very high computational resources and training time,
Spanhol et al. (2016a)	Deep Learning	85.6 ± 4.8	---	Benign/Malignant	Public, BreakHis	• Needs to improve performance, • Sn, Fm, and MCC need to be reported, • It requires very high computational resources and training time.
Janowczyk and Madabhushi (2016)	Deep Learning	84.68	---	Invasive/ In situ	Exclusive, Not given	• Needs to improve performance, • Sn, Fm, and MCC need to be reported, • Requires very high computational resources and training time, • The use of a public dataset can show different results.
Huang and Chung (2018)	Deep Learning	98.5	---	Carcinoma/ Non-carcinoma	Public, BCBH	• Sn, Fm, and MCC need to be reported, • Requires very high computational resources and training time, • Need to be tested using a larger size dataset.

Reference	**Classification approach**	**Ac (%)**	**Sn (%)**	**Class labels**	**Dataset type, name**	**Limitations**
Araujo et al. (2017)	Deep Learning	Initial = 90, Extended = 81.3	Initial = (In situ = 80, Invasive = 100), Extended = (In situ= 50, Invasive = 100)	In situ/Invasive	Public, BCBH	• Needs to improve performance, • Requires high computational resources and training time, • Need to be tested using a larger size dataset.
Proposed	Traditional Machine Learning (GK, kNN)	Initial = 90, Extended = 93	Initial = (In situ = 90, Invasive = 90), Extended = (In situ= 100, Invasive = 87.50)	In situ (i.e., Non-carcinoma) / Invasive (i.e., Carcinoma	Public, BCBH	• Need to be tested using a larger size dataset.

6. Conclusion

An automated breast cancer detection is performed by classifying histology images in carcinoma and non-carcinoma cancer types. A publicly available BCBH dataset of DH images was utilized to develop the BC detection model. Next, a few images pre-processing tasks like stain normalization, training set augmentation, and rescaling is performed to enhance the BC detection performance. Where stain normalization harmonizes the inherent color inconsistencies of DH images and image augmentation is required to create synthetic sample data to improve classifiers' detection performance. Moreover, rescaling is adopted to extract features efficiently from smaller size images. The preprocessed images are used to extract both local and global types of features using ten feature extraction methods. An MFT is created for each feature extraction method using an augmented training set, initial testing set, and extended testing set. Afterwards, the features of all MFTs are sorted by their weights calculated through the NCA method to create SMFTs. An algorithm is developed to select the minimum number of features from SMFTs to train seven heterogeneous ML classifiers. It has been observed from numerous experiments that the performance of ML classifiers using SMFTs is fluctuating up to59features and then started degrading as the weight is decreasing. Furthermore, the goal of the feature selection algorithm is to look for the best performing feature extraction method (out of ten), the best ML classifier (out of seven) using the minimum number of features. After performing many experiments, it has been found that for the initial testing set GK outperformed the other ML classifiers using only 51(out of 64) Harris features by showing 90% results for each like Ac, Sn, and Fm with 0.8 MCC value. On the other hand, while using extended testing kNN outperformed the rest of the classifiers using only 3 (out of 64) Harris features by achieving 93% Ac, 94% Fm, and 0.8819 MCC value. Thus, it can be concluded that Harris features outperformed all feature extraction methods. The performance of the proposed BC detection model shows that the model is efficient (consumes less training and detection time), cost-effective (can be trained on a normal desktop computer), and feasible (requires fewer number of labelled images to show good results) solution for BC detection using DH images. Thus, the model can be implemented efficiently in any health care centre to assist doctors as a second opinion for BC early detection using DH images, especially in the low privileged area of a country. As future work, the proposed model can be tested by using other datasets with a large number of BC DH images. Moreover, other types of texture features and the wavelet-based statistical features can be extracted to make further analysis.

Acknowledgment: This work was supported by University Malaya Research Grant Program – [grant number GPF016D-2019]

AUTHOR CONTRIBUTION*:* Ghulam Murtaza(GM), Ainuddin Wahid Abdul Wahab(AWAW), Ghulam Raza(GR), Liyana Shuib(LS)

Data curation: GM, LS.
Methodology: GM, AWAW.
Software: GM.
Supervision: AWAW, LS.
Writing – original draft: GM.
Writing – review & editing: GM, AWAW, GR, LS.
Funding: AWAW.

DATA AVAILABILTY STATEMENT: A publicly available dataset is used for this research work.

URL: https://rdm.inesctec.pt/dataset/nis-2017-003.

CONFLICT OF INTEREST: None

FUNDING: This work was supported by University Malaya Research Grant Program – [grant number GPF016D-2019]

References

[1] Alcantarilla, P. F., Bartoli, A., & Davison, A. J. (2012). KAZE features. Paper presented at the European Conference on Computer Vision.

[2] American Cancer Society. (2018). Breast Cancer Facts & Figures 2017-2018. Atlanta: American Cancer Society, Inc. 2017. Retrieved 03/08/2019

[3] Araujo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., . . . Campilho, A. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. Plos One, 12(6), 14. doi: 10.1371/journal.pone.0177544

[4] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer vision and image understanding, 110(3), 346-359.

[5] Belsare, A. D., Mushrif, M. M., Pangarkar, M. A., & Meshram, N. (2015). Classification of Breast Cancer Histopathology Images using Texture Feature Analysis Tencon 2015 - 2015 Ieee Region 10 Conference. New York: Ieee.

[6] Bige, O., Demir, A., Saygili, U., Gode, F., Uslu, T., & Koyuncuoglu, M. (2011). Frozen section diagnoses of 578 ovarian tumors made by pathologists with and without expertise on gynecologic pathology. Gynecologic Oncology, 123(1), 43-46. doi: https://doi.org/10.1016/j.ygyno.2011.06.030

[7] Biswas, R., Nath, A., & Roy, S. (2016). Mammogram Classification using Gray-Level Cooccurrence Matrix for Diagnosis of Breast Cancer. New York: Ieee.

[8] BREASTCANCER.ORG. (2015, October 7, 2015). Biopsy techniques. from https://www.breastcancer.org/symptoms/testing/types/biopsy

[9] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection.

[10] Desjardins, A. U. (1960). Is the Pathologist Infallible? JAMA Internal Medicine, 106(5), 596-602. doi: 10.1001/archinte.1960.03820050008003

[11] Gupta, V., & Bhavsar, A. (2018). Sequential modeling of deep features for breast cancer histopathological image classification. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[12] Gupta, V., Bhavsar, A., & Ieee. (2017). Breast Cancer Histopathological Image Classification: Is Magnification Important? 2017 Ieee Conference on Computer Vision and Pattern Recognition Workshops (pp. 769-776). New York: Ieee.

[13] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: a review. IEEE reviews in biomedical engineering, 2, 147-171. doi: 10.1109/RBME.2009.2034865

[14] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. Scientific Reports, 7(1). doi: 10.1038/s41598-017-04075-z

[15] Harris, C. G., & Stephens, M. (1988). A combined corner and edge detector. Paper presented at the Alvey vision conference.

[16] Hassanien, A. E., & Ali, J. M. (2006). Rough set approach for classification of breast cancer mammogram images. In V. DiGesu, F. Masulli & A. Petrosino (Eds.), Fuzzy Logic and Applications (Vol. 2955, pp. 224-231). Berlin: Springer-Verlag Berlin.

[17] Huang, Y., & Chung, A. C.-S. (2018, 2018//). Improving High Resolution Histology Image Classification with Deep Spatial Fusion Network. Paper presented at the Computational Pathology and Ophthalmic Medical Image Analysis, Cham.

[18] Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics, 7(1), 29-29. doi: 10.4103/2153-3539.186902

[19] Jianbo, S., & Tomasi, C. (1994). Good features to track. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[20] Khan, A. M., Rajpoot, N., Treanor, D., & Magee, D. (2014). A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Transactions on Biomedical Engineering, 61(6), 1729-1738.

[21] Komura, D., & Ishikawa, S. (2018). Machine Learning Methods for

Histopathological Image Analysis. Computational and Structural Biotechnology Journal, 16, 34-42. doi: https://doi.org/10.1016/j.csbj.2018.01.001

[22] Kontos, K., & Maragoudakis, M. (2013). Breast Cancer Detection in Mammogram Medical Images with Data Mining Techniques. In H. Papadopoulos, A. S. Andreou, L. Iliadis & I. Maglogiannis (Eds.), Artificial Intelligence Applications and Innovations, Aiai 2013 (Vol. 412, pp. 336-347). Berlin: Springer-Verlag Berlin.

[23] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Paper presented at the Advances in neural information processing systems.

[24] Leutenegger, S., Chli, M., & Siegwart, R. (2011). BRISK: Binary robust invariant scalable keypoints. Paper presented at the 2011 IEEE international conference on computer vision (ICCV).

[25] Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., . . . Thomas, N. E. (2009). A method for normalizing histology slides for quantitative analysis. Paper presented at the Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on.

[26] Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing, 22(10), 761-767. doi: https://doi.org/10.1016/j.imavis.2004.02.006

[27] McCann, M. T., Ozolek, J. A., Castro, C. A., Parvin, B., & Kovacevic, J. (2015). Automated histology analysis: Opportunities for signal processing. IEEE Signal Processing Magazine, 32(1), 78-87.

[28] Moravec, H. P. (1980). Obstacle avoidance and navigation in the real world by a seeing robot rover: Stanford Univ CA Dept of Computer Science.

[29] Murtaza, G., Shuib, L., Abdul Wahab, A. W., Mujtaba, G., Mujtaba, G., Nweke, H. F., . . . Azmi, N. A. (2020). Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artificial Intelligence Review, 53(3), 1655-1720. doi: 10.1007/s10462-019-09716-5

[30] Murtaza, G., Shuib, L., Mujtaba, G., & Raza, G. (2019). Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach. Multimedia Tools and Applications. doi: 10.1007/s11042-019-7525-4

[31] Murtaza, G., Shuib, L., Wahab, A. W. A., Mujtaba, G., Raza, G., & Azmi, N. A. (2019). Breast cancer classification using digital biopsy histopathology images through transfer learning. Paper presented at the Journal of Physics: Conference Series.

[32] Murtaza, G., Shuib, L., Wahab, A. W. A., Mujtaba, G., Raza, G. J. M. T., & Applications. (2020). Ensembled deep convolution neural network-based breast cancer classification with misclassification reduction algorithms. 1-33.

[33] Nusantara, A. C., Purwanti, E., & Soelistiono, S. (2016). CLASSIFICATION OF DIGITAL MAMMOGRAM BASED ON NEAREST-NEIGHBOR METHOD FOR BREAST CANCER DETECTION. International Journal of Technology, 7(1), 71-77. doi: 10.14716/ijtech.v7i1.1393

[34] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence(7), 971-987.

[35] Ponraj, N., Poongodi, & Mercy, M. (2017). Texture Analysis of Mammogram for the Detection of

Breast Cancer using LBP and LGP: A Comparison 2016 Eighth International Conference on Advanced Computing (pp. 182-185). New York: Ieee.

[36] Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. IEEE Computer graphics and applications, 21(5), 34-41.

[37] Rosten, E., & Drummond, T. (2005). Fusing points and lines for high performance tracking. Paper presented at the ICCV.

[38] Rubin, R., Strayer, D. S., & Rubin, E. (2008). Rubin's pathology: clinicopathologic foundations of medicine: Lippincott Williams & Wilkins.

[39] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. R. (2011). ORB: An efficient alternative to SIFT or SURF. Paper presented at the ICCV.

[40] Song, Y., Zou, J. J., Chang, H., & Cai, W. (2017). Adapting fisher vectors for histopathology image classification. Paper presented at the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).

[41] Spanhol, F. A., Oliveira, L. S., Cavalin, P. R., Petitjean, C., & Heutte, L. (2017). Deep features for breast cancer histopathological image classification. Paper presented at the Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on.

[42] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016a). Breast cancer histopathological image classification using convolutional neural networks. Paper presented at the Neural Networks (IJCNN), 2016 International Joint Conference on.

[43] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016b). A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 63(7), 1455-1462.

[44] Tabar, L., Vitak, B., Chen, T. H., Yen, A. M., Cohen, A., Tot, T., . . . Duffy, S. W. (2011). Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology, 260(3), 658-663. doi: 10.1148/radiol.11110469

[45] Teresa Araújo, G. A., Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, Aurélio Campilho. (2015, November 15, 2018, 7:19 PM (UTC+08:00)). Bioimaging Challenge 2015 Breast Histology Dataset. from https://rdm.inesctec.pt/dataset/nis-2017-003

[46] Wan, T., Cao, J., Chen, J., & Qin, Z. (2017). Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. Neurocomputing, 229, 34-44. doi: https://doi.org/10.1016/j.neucom.2016.05.084

[47] Wan, T., Liu, X., Chen, J., & Qin, Z. (2014, 27-30 Oct. 2014). Wavelet-based statistical features for distinguishing mitotic and non-mitotic cells in breast cancer histopathology. Paper presented at the 2014 IEEE International Conference on Image Processing (ICIP).

[48] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718.

[49] Weickert, J., Ishikawa, S., & Imiya, A. (1999). Linear scale-space has first been proposed in Japan. Journal of Mathematical Imaging and Vision, 10(3), 237-252.

[50] Zheng, Y., Jiang, Z., Xie, F., Zhang, H., Ma, Y., Shi, H., & Zhao, Y. (2017). Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. Pattern Recognition, 71, 14-25. doi: 10.1016/j.patcog.2017.05.010

Appendix-A

Table 1: Other Performance Evaluation Metrics of using the original image

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Sp**		70.00	20.00	100.00	60.00	10.00	40.00	70.00	62.50	12.50	100.00	62.50	12.50	62.50	37.50
	Pr		72.73	55.56	NaN	66.67	52.63	45.46	62.50	66.67	50.00	NaN	62.50	53.33	62.50	44.44
	Confusion Matrix	**TP**	8	10	0	8	10	5	5	6	7	0	5	8	5	4
		FN	2	0	10	2	0	5	5	2	1	8	3	0	3	4
		FP	3	8	0	4	9	6	3	3	7	0	3	7	3	5
		TN	7	2	10	6	1	4	7	5	1	8	5	1	5	3
MinEigen	**Sp**		80.00	90.00	50.00	60.00	50.00	70.00	20.00	37.50	62.50	75.00	62.50	37.50	62.50	75.00
	Pr		66.67	66.67	37.50	33.33	50.00	70.00	38.46	28.57	40.00	50.00	57.14	44.44	25.00	71.43
	Confusion Matrix	**TP**	4	2	3	2	5	7	5	2	2	2	4	4	1	5
		FN	6	8	7	8	5	3	5	6	6	6	4	4	7	3
		FP	2	1	5	4	5	3	8	5	3	2	3	5	3	2
		TN	8	9	5	6	5	7	2	3	5	6	5	3	5	6
Harris	**Sp**		40.00	90.00	80.00	80.00	60.00	40.00	40.00	50.00	50.00	37.50	37.50	50.00	75.00	62.50
	Pr		40.00	75.00	66.67	66.67	55.56	53.85	45.46	60.00	50.00	28.57	37.50	42.86	66.67	40.00
	Confusion Matrix	**TP**	4	3	4	4	5	7	5	6	4	2	3	3	4	2
		FN	6	7	6	6	5	3	5	2	4	6	5	5	4	6
		FP	6	1	2	2	4	6	6	4	4	5	5	4	2	3
		TN	4	9	8	8	6	4	4	4	4	3	3	4	6	5
LBP	**Sp**		50.00	50.00	50.00	40.00	80.00	40.00	50.00	75.00	87.50	75.00	62.50	62.50	87.50	50.00
	Pr		61.54	58.33	64.29	57.14	77.78	57.14	61.54	66.67	80.00	66.67	66.67	62.50	80.00	33.33
	Confusion Matrix	**TP**	8	7	9	8	7	8	8	4	4	4	6	5	4	2
		FN	2	3	1	2	3	2	2	4	4	4	2	3	4	6
		FP	5	5	5	6	2	6	5	2	1	2	3	3	1	4
		TN	5	5	5	4	8	4	5	6	7	6	5	5	7	4
KAZ	**Sp**		90.00	90.00	90.00	50.00	80.00	70.00	50.00	100.00	87.50	100.00	62.50	50.00	100.00	37.50

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	Pr		66.67	66.67	66.67	50.00	71.43	40.00	50.00	100.00	66.67	100.00	40.00	42.86	100.00	44.44
	Confusion Matrix	**TP**	2	2	2	5	5	2	5	1	2	1	2	3	1	4
		FN	8	8	8	5	5	8	5	7	6	7	6	5	7	4
		FP	1	1	1	5	2	3	5	0	1	0	3	4	0	5
		TN	9	9	9	5	8	7	5	8	7	8	5	4	8	3
MSER	**Sp**		90.00	50.00	90.00	90.00	80.00	100.00	80.00	87.50	75.00	87.50	100.00	75.00	87.50	87.50
	Pr		66.67	64.29	80.00	75.00	60.00	100.00	66.67	75.00	71.43	75.00	100.00	75.00	66.67	66.67
	Confusion Matrix	**TP**	2	9	4	3	3	4	4	3	5	3	5	6	2	2
		FN	8	1	6	7	7	6	6	5	3	5	3	2	6	6
		FP	1	5	1	1	2	0	2	1	2	1	0	2	1	1
		TN	9	5	9	9	8	10	8	7	6	7	8	6	7	7
SURF	**Sp**		40.00	70.00	80.00	90.00	100.00	50.00	90.00	62.50	100.00	100.00	75.00	75.00	62.50	75.00
	Pr		57.14	0.00	0.00	0.00	100.00	61.54	50.00	57.14	100.00	NaN	33.33	60.00	57.14	60.00
	Confusion Matrix	**TP**	8	0	0	0	1	8	1	4	1	0	1	3	4	3
		FN	2	10	10	10	9	2	9	4	7	8	7	5	4	5
		FP	6	3	2	1	0	5	1	3	0	0	2	2	3	2
		TN	4	7	8	9	10	5	9	5	8	8	6	6	5	6
FAST	**Sp**		50.00	100.00	100.00	100.00	90.00	20.00	80.00	25.00	62.50	75.00	62.50	62.50	50.00	62.50
	Pr		64.286	100.00	100.00	100.00	50.00	46.67	50.00	57.143	62.50	50.00	50.00	62.50	55.56	40.00
	Confusion Matrix	**TP**	9	2	2	2	1	7	2	8	5	2	3	5	5	2
		FN	1	8	8	8	9	3	8	0	3	6	5	3	3	6
		FP	5	0	0	0	1	8	2	6	3	2	3	3	4	3
		TN	5	10	10	10	9	2	8	2	5	6	5	5	4	5
BRISK	**Sp**		50.00	90.00	80.00	80.00	90.00	60.00	80.00	12.50	50.00	87.50	75.00	62.50	37.50	75.00
	Pr		61.54	50.00	50.00	50.00	66.67	60.00	33.33	50.00	60.00	66.67	71.43	62.50	44.44	71.43
	Confusion	**TP**	8	1	2	2	2	6	1	7	6	2	5	5	4	5
		FN	2	9	8	8	8	4	9	1	2	6	3	3	4	3

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
		FP	5	1	2	2	1	4	2	7	4	1	2	3	5	2
		TN	5	9	8	8	9	6	8	1	4	7	6	5	3	6
ORB	**Sp**		50.00	40.00	70.00	60.00	60.00	60.00	30.00	50.00	50.00	50.00	50.00	50.00	12.50	62.50
	Pr		50.00	50.00	57.14	50.00	69.23	55.56	41.67	60.00	42.86	50.00	50.00	50.00	41.67	50.00
	Confusion Matrix	**TP**	5	6	4	4	9	5	5	6	3	4	4	4	5	3
		FN	5	4	6	6	1	5	5	2	5	4	4	4	3	5
		FP	5	6	3	4	4	4	7	4	4	4	4	4	7	3
		TN	5	4	7	6	6	6	3	4	4	4	4	4	1	5

Table 2: Other Performance Evaluation Metrics of using image augmentation algorithm

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Sp**		40.00	60.00	40.00	90.00	60.00	50.00	70.00	75.00	25.00	75.00	62.50	75.00	87.50	37.50
	Pr		60.00	69.23	62.50	87.50	69.23	64.29	72.73	77.78	57.14	60.00	70.00	75.00	83.33	44.44
	Confusion Matrix	**TP**	9	9	10	7	9	9	8	7	8	3	7	6	5	4
		FN	1	1	0	3	1	1	2	1	0	5	1	2	3	4
		FP	6	4	6	1	4	5	3	2	6	2	3	2	1	5
		TN	4	6	4	9	6	5	7	6	2	6	5	6	7	3
MinEigen	**Sp**		30.00	40.00	90.00	70.00	100.00	60.00	60.00	0.00	50.00	12.50	87.50	87.50	62.50	75.00
	Pr		58.82	62.50	87.50	75.00	100.00	63.64	66.67	50.00	60.00	50.00	87.50	85.71	70.00	66.67
	Confusion Matrix	**TP**	10	10	7	9	7	7	8	8	6	7	7	6	7	4
		FN	0	0	3	1	3	3	2	0	2	1	1	2	1	4
		FP	7	6	1	3	0	4	4	8	4	7	1	1	3	2
		TN	3	4	9	7	10	6	6	0	4	1	7	7	5	6
Harris	**Sp**		50.00	30.00	50.00	50.00	40.00	60.00	70.00	37.50	62.50	37.50	87.50	75.00	37.50	37.50
	Pr		66.67	56.25	61.54	61.54	60.00	69.23	70.00	61.54	72.73	50.00	83.33	77.78	58.33	37.50
	Confusion	**TP**	10	9	8	8	9	9	7	8	8	5	5	7	7	3
		FN	0	1	2	2	1	1	3	0	0	3	3	1	1	5

			Initial Testing Set							**Extended Testing Set**						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
		FP	5	7	5	5	6	4	3	5	3	5	1	2	5	5
		TN	5	3	5	5	4	6	7	3	5	3	7	6	3	3
LBP	**Sp**		30.00	50.00	50.00	50.00	70.00	50.00	50.00	37.50	37.50	37.50	75.00	62.50	37.50	50.00
	Pr		58.82	64.29	64.29	64.29	75.00	64.29	66.67	61.54	58.33	44.44	77.78	72.73	61.54	60.00
	Confusion Matrix	**TP**	10	9	9	9	9	9	10	8	7	4	7	8	8	6
		FN	0	1	1	1	1	1	0	0	1	4	1	0	0	2
		FP	7	5	5	5	3	5	5	5	5	5	2	3	5	4
		TN	3	5	5	5	7	5	5	3	3	3	6	5	3	4
KAZE	**Sp**		40.00	70.00	80.00	70.00	90.00	90.00	70.00	75.00	62.50	0.00	75.00	37.50	62.50	0.00
	Pr		62.50	72.73	80.00	76.92	87.50	87.50	72.73	75.00	66.67	50.00	75.00	61.54	66.67	50.00
	Confusion Matrix	**TP**	10	8	8	10	7	7	8	6	6	8	6	8	6	8
		FN	0	2	2	0	3	3	2	2	2	0	2	0	2	0
		FP	6	3	2	3	1	1	3	2	3	8	2	5	3	8
		TN	4	7	8	7	9	9	7	6	5	0	6	3	5	0
MSER	**Sp**		50.00	90.00	70.00	80.00	70.00	90.00	80.00	75.00	50.00	62.50	100.00	87.50	62.50	12.50
	Pr		61.54	88.89	70.00	77.78	70.00	87.50	81.82	71.43	63.64	50.00	100.00	87.50	70.00	22.22
	Confusion Matrix	**TP**	8	8	7	7	7	7	9	5	7	3	5	7	7	2
		FN	2	2	3	3	3	3	1	3	1	5	3	1	1	6
		FP	5	1	3	2	3	1	2	2	4	3	0	1	3	7
		TN	5	9	7	8	7	9	8	6	4	5	8	7	5	1
SURF	**Sp**		60.00	60.00	70.00	60.00	60.00	70.00	80.00	25.00	37.50	100.00	62.50	62.50	37.50	100.00
	Pr		69.23	69.23	72.73	69.23	66.67	75.00	75.00	53.85	58.33	NaN	66.67	70.00	58.33	NaN
	Confusion Matrix	**TP**	9	9	8	9	8	9	6	7	7	0	6	7	7	0
		FN	1	1	2	1	2	1	4	1	1	8	2	1	1	8
		FP	4	4	3	4	4	3	2	6	5	0	3	3	5	0
		TN	6	6	7	6	6	7	8	2	3	8	5	5	3	8
FAST	**Sp**		60.00	100.00	100.00	100.00	90.00	70.00	100.00	50.00	75.00	100.00	87.50	75.00	50.00	100.00
	Pr		71.43	100.00	100.00	100.00	80.00	72.73	100.00	66.667	77.78	NaN	80.00	75.00	63.64	NaN

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	Confusion Matrix	**TP**	10	3	2	2	4	8	4	8	7	0	4	6	7	0
		FN	0	7	8	8	6	2	6	0	1	8	4	2	1	8
		FP	4	0	0	0	1	3	0	4	2	0	1	2	4	0
		TN	6	10	10	10	9	7	10	4	6	8	7	6	4	8
BRISK	**Sp**		50.00	75.00	100.00	87.50	87.50	50.00	100.00	80.00	60.00	90.00	100.00	70.00	60.00	100.00
	Pr		66.67	75.00	NaN	85.71	83.33	63.64	NaN	83.33	50.00	75.00	100.00	57.14	71.43	100.00
	Confusion Matrix	**TP**	8	6	0	6	5	7	0	10	4	3	3	4	10	3
		FN	0	2	8	2	3	1	8	0	6	7	7	6	0	7
		FP	4	2	0	1	1	4	0	2	4	1	0	3	4	0
		TN	4	6	8	7	7	4	8	8	6	9	10	7	6	10
ORB	**Sp**		90.00	60.00	40.00	40.00	50.00	60.00	50.00	62.50	62.50	50.00	87.50	62.50	62.50	37.50
	Pr		88.89	63.64	57.14	57.14	64.29	66.67	61.54	62.50	66.67	42.86	85.71	66.67	66.67	44.44
	Confusion Matrix	**TP**	8	7	8	8	9	8	8	5	6	3	6	6	6	4
		FN	2	3	2	2	1	2	2	3	2	5	2	2	2	4
		FP	1	4	6	6	5	4	5	3	3	4	1	3	3	5
		TN	9	6	4	4	5	6	5	5	5	4	7	5	5	3

Table 3: Other Performance Evaluation Metrics of using image augmentation with feature selection algorithm

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
HOG	**Sp**		90.00	50.00	70.00	70.00	70.00	70.00	90.00	50.00	62.50	62.50	75.00	100.00	75.00	87.50
	Pr		90.00	66.67	76.92	76.92	76.92	75.00	88.89	66.67	72.73	72.73	77.78	100.00	80.00	87.50
	Confusion Matrix	**TP**	9	10	10	10	10	9	8	8	8	8	7	7	8	7
		FN	1	0	0	0	0	1	2	0	0	0	1	1	0	1
		FP	1	5	3	3	3	3	1	4	3	3	2	0	2	1
		TN	9	5	7	7	7	7	9	4	5	5	6	8	6	7
MinEigen	**Sp**		60.00	50.00	90.00	70.00	100.00	90.00	80.00	62.50	62.50	75.00	100.00	100.00	100.00	87.50
	Pr		71.43	66.67	87.50	75.00	100.00	88.89	80.00	72.73	70.00	80.00	100.00	100.00	100.00	87.50

			Initial Testing Set							Extended Testing Set						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
	Confusion Matrix	**TP**	10	10	7	9	8	8	8	8	7	8	6	7	7	7
		FN	0	0	3	1	2	2	2	0	1	0	2	1	1	1
		FP	4	5	1	3	0	1	2	3	3	2	0	0	0	1
		TN	6	5	9	7	10	9	8	5	5	6	8	8	8	7
Harris	**Sp**		70.00	40.00	70.00	60.00	60.00	90.00	70.00	75.00	62.50	62.50	62.50	87.50	75.00	87.50
	Pr		76.92	62.50	75.00	69.23	71.43	90.00	75.00	80.00	72.73	72.73	72.73	88.89	80.00	87.50
	Confusion Matrix	**TP**	10	10	9	9	10	9	9	8	8	8	8	8	8	7
		FN	0	0	1	1	0	1	1	0	0	0	0	0	0	1
		FP	3	6	3	4	4	1	3	2	3	3	3	1	2	1
		TN	7	4	7	6	6	9	7	6	5	5	5	7	6	7
LBP	**Sp**		50.00	40.00	60.00	60.00	90.00	50.00	60.00	50.00	62.50	62.50	87.50	62.50	62.50	100.00
	Pr		66.67	62.50	71.43	69.23	88.89	66.67	71.43	66.67	70.00	70.00	87.50	72.73	72.73	100.00
	Confusion Matrix	**TP**	10	10	10	9	8	10	10	8	7	7	7	8	8	6
		FN	0	0	0	1	2	0	0	0	1	1	1	0	0	2
		FP	5	6	4	4	1	5	4	4	3	3	1	3	3	0
		TN	5	4	6	6	9	5	6	4	5	5	7	5	5	8
KAZE	**Sp**		60.00	80.00	90.00	80.00	90.00	70.00	80.00	75.00	87.50	75.00	62.50	75.00	75.00	75.00
	Pr		71.43	81.82	90.00	83.33	90.00	76.92	81.82	77.78	85.71	77.78	72.73	80.00	75.00	80.00
	Confusion Matrix	**TP**	10	9	9	10	9	10	9	7	6	7	8	8	6	8
		FN	0	1	1	0	1	0	1	1	2	1	0	0	2	0
		FP	4	2	1	2	1	3	2	2	1	2	3	2	2	2
		TN	6	8	9	8	9	7	8	6	7	6	5	6	6	6
MSER	**Sp**		70.00	80.00	70.00	80.00	80.00	100.00	80.00	62.50	50.00	87.50	87.50	87.50	50.00	75.00
	Pr		75.00	81.82	75.00	80.00	80.00	100.00	81.82	66.67	63.64	83.33	85.71	87.50	66.67	80.00
	Confusion	**TP**	9	9	9	8	8	8	9	6	7	5	6	7	8	8
		FN	1	1	1	2	2	2	1	2	1	3	2	1	0	0
		FP	3	2	3	2	2	0	2	3	4	1	1	1	4	2

			Initial Testing Set							**Extended Testing Set**						
			SVM	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**	**SVM**	**NB**	**LR**	**LDA**	**kNN**	**GK**	**DT**
		TN	7	8	7	8	8	10	8	5	4	7	7	7	4	6
SURF	**Sp**		50.00	50.00	70.00	80.00	70.00	60.00	80.00	62.50	50.00	50.00	50.00	100.00	50.00	75.00
	Pr		66.67	66.67	75.00	81.82	75.00	69.23	80.00	70.00	63.64	63.64	63.64	100.00	63.64	77.78
	Confusion Matrix	**TP**	10	10	9	9	9	9	8	7	7	7	7	6	7	7
		FN	0	0	1	1	1	1	2	1	1	1	1	2	1	1
		FP	5	5	3	2	3	4	2	3	4	4	4	0	4	2
		TN	5	5	7	8	7	6	8	5	4	4	4	8	4	6
FAST	**Sp**		70.00	80.00	100.00	100.00	100.00	60.00	100.00	50.00	87.50	100.00	100.00	100.00	75.00	87.50
	Pr		76.923	66.67	100.00	100.00	100.00	71.43	100.00	66.667	87.50	100.00	100.00	100.00	80.00	85.71
	Confusion Matrix	**TP**	10	4	3	3	4	10	4	8	7	4	5	5	8	6
		FN	0	6	7	7	6	0	6	0	1	4	3	3	0	2
		FP	3	2	0	0	0	4	0	4	1	0	0	0	2	1
		TN	7	8	10	10	10	6	10	4	7	8	8	8	6	7
BRISK	**Sp**		80.00	90.00	100.00	100.00	100.00	70.00	90.00	50.00	75.00	100.00	87.50	100.00	62.50	87.50
	Pr		83.33	80.00	100.00	100.00	100.00	76.92	80.00	66.67	75.00	100.00	85.71	100.00	72.73	87.50
	Confusion Matrix	**TP**	10	4	3	3	4	10	4	8	6	3	6	6	8	7
		FN	0	6	7	7	6	0	6	0	2	5	2	2	0	1
		FP	2	1	0	0	0	3	1	4	2	0	1	0	3	1
		TN	8	9	10	10	10	7	9	4	6	8	7	8	5	7
ORB	**Sp**		70.00	60.00	80.00	70.00	80.00	80.00	80.00	62.50	75.00	62.50	50.00	62.50	100.00	100.00
	Pr		75.00	69.23	81.82	76.92	81.82	80.00	80.00	72.73	80.00	72.73	66.67	72.73	100.00	100.00
	Confusion Matrix	**TP**	9	9	9	10	9	8	8	8	8	8	8	8	6	6
		FN	1	1	1	0	1	2	2	0	0	0	0	0	2	2
		FP	3	4	2	3	2	2	2	3	2	3	4	3	0	0
		TN	7	6	8	7	8	8	8	5	6	5	4	5	8	8

Table 4: List of abbreviations

Abbreviation	Full-Form
Ac	Accuracy
BC	Breast cancer
BCBH	Bioimaging challenge 2015 breast histology
BreakHis	Breast cancer histopathological image classification
BRISK	Binary robust invariant scalable keypoint
BT	Breast tumor
DH	Digital histology
DT	Decision tree
FAST	Features from accelerated segment test
FEMs	Feature extraction methods
Fm	F-measure
HoG	Histogram-oriented gradients
kNN	kNearest Neighbors
LBP	Local binary patterns
LR	Linear regression
MFT	Master feature table
MG	Mammogram
MinEigen	Minimum eigenvalue
ML	Machine learning
MRI	Magnetic resonance imaging
MSER	Maximally stable extremal regions
NB	Naive Bayes
ORB	Oriented fast and rotated brief
PEMs	Performance evaluation metrics
SMFT	Sorted master feature table
Sn	Sensitivity
Sp	Specificity
SURF	Speed-up robust feature
SVM	Support vector machine
US	Ultrasound
WSI	Whole slide image

Vol. 5, No. 1 | January – June 2021

Impact Assessment of Urban Pull-factors on Uncontrolled Urbanization: Evidence from Pakistan

Hassam Bin Waseem[1], Mir Aftab Hussain Talpur[1]

Abstract:
Uncontrolled urbanization is a global phenomenon, currently sweeping through developing countries like Pakistan. Being the 6th most populous country, its strategic urban locations receive a humongous migrant influx. A prime victim is Hyderabad, Pakistan, the Sindh's second-largest urban settlement after Karachi that experiences enormous urban problems due to immense urbanization. Hence, the study aims to propose policy recommendations to curtail urbanization rate of Hyderabad, Sindh and Pakistan. The objective of this study is to measure the impacts of pull-factors in perception of migrants so that the policy recommendations could be proposed on most responsible factors. For this study, 3 urban Talukas (a district's subdivision) of district Hyderabad were selected. While a sample of 400 migrants was procured using purposive and snowball sampling techniques for the questionnaire survey. Utilizing a large survey data, the results were obtained using descriptive statistics and multiple regression. Results showed a significant impact of economic and socio-cultural pull-factors like better employment opportunities (0.152), higher income probability (0.222), job security (0.779), easy access to facilities (0.763), better outlook and hope for the future (0.324), and comfortable and diverse lifestyle (0.159) in causing immense urbanization. Nevertheless, desire for independence and access to basic social services appeared to have a negative impact. The proposed policy recommendations on significant pull-factors could be considered as a first step to handle uncontrolled urbanization. Moreover, the study is significant as it particularly contributes to apprehend Hyderabad's various interrelated urban issues. In general terms, this research can be considered as a role model for high-density third world cities suffering from relatable urban malaise.

Keywords: *Uncontrolled Urbanization; Pull-factors; Multiple Regression; Third World Cities; Policy Recommendations*

1. Introduction

Uncontrolled urbanization is a global urban problem, which can be defined as, "the process of rural to urban migration complementing the immense and rapid growth of urban areas" [1]. Hypothesize as a consequence of population increase, industrialization, technology reforms, infrastructure services, and fiscal advancements; it specifically leads to urban transformations [2, 3]. Being the most vital element of internal migration and hasty urban growth; it is an influential sign of regional disparity concerning economy and lifestyle [4, 5]. Intrinsically, the process involves urban development activities as poverty-driven rural laborers satisfy the huge labor requirement of urban centers [6].

Urbanization is a multi-dimensional event that occurred for numerous whys and

[1] Mehran University of Engineering & Technology, Jamshoro, Sindh, Pakistan
Corresponding Author: hassam.waseem@gmail.com

wherefores. Its causes and factors may differ between the countries and inside a country too. Irrespective of its several stages such as suburbanization and ex-urbanization, the phenomenon leads to multifaceted urban spatial structure, declining agricultural land, deviating employment and residential locations, and pressures on urban and rural livelihoods [7-10]. Mostly, the phenomenon relies on four factors, i.e. push, pull, stay, and return; of which, pull-factors are the most prominent determinants [11, 12]. They convince a person to attempt migration in search of better economic and socio-cultural conditions for him/herself and/or for their family [3]. Furthermore, several phenomena of urban areas like mono-centric, multi-centric, dwindling and flourishing, refine and shape their decisions [13]. Complementing this, the variances in migration rates directed discrepancies in physical and societal patterns in the city centers [14]. Thus, having profound impacts on urban areas, its linkages, and overall urban expansion.

In the context of third world countries, the phenomenon of urbanization is conceived as an illimitable concern that results in strained health and education budgets, and complicated unemployment level reduction [15, 16]. Though the increase in urbanization aids in terms of poverty reduction and economic growth [10, 16, 17]. Nonetheless, the snags are coupled with urban issues, like ever-expanding slums, scarce basic urban services, inadequate housing, overpopulation, ill-planning of shelters and industries along with nonconformity of land use bylaws/codes/standards [18]. Furthermore, the phenomenon faded urban-suburban rift that led to urban sprawl [19].

The urban planner's perspective in this dilemma is the exaggeration of land competition on urbanization focal points either through land development or redevelopment. They are concerned as its consequences would affect third world countries by soaking up physical, economic, and socio-cultural transformations in the form of urban activities [20]. The intrusion of valuable peri-urban land is the climax of this paradigm, i.e. an issue affecting sustainability [21]. This dilemma is termed as 'Desakota' referring to the concentrated blend among agricultural and non-agricultural land uses on urban fringes [22]. The migration within third world countries (such as, Pakistan) is usually grounded on pull-factors that raise questions regarding the linkages between urbanization and several pull-factors [23]. This objective-oriented migration makes urbanization as Pakistan's recent developmental issue, for which, no migration profile is formulated [3, 24].

The immense urbanization pace is on the verge these days in Pakistan. It would continue at a faster pace for upcoming years as Pakistan's economy is escalating and reforming [25]. Surprisingly, no one portrayed pull-factors, such as rural environmental degradation, and so on, as its major drivers [26]. Preliminary results of the Pakistan Census 2017 declared one of the provinces of Pakistan, i.e. Sindh as the most urbanized province with 52.02% urban population, and Hyderabad as the second most urbanized district, after Karachi with 1,834,371 urbanites [27]. As mentioned earlier, its strategic location attracts more populace flow making it a victim of urban agglomeration. This worst scenario of uncontrolled urbanization was also highlighted in the Hyderabad Master Plan (2007-2020). The master plan document clearly declared that urban pull is causing mushroom growth complemented with urban issues, like over-stretched and imbalanced infrastructure, over-burdened facilities, bad condition of roads, traffic problems, and so on [28]. Moreover, the extreme flow of migrants towards Hyderabad validates a strong urban pull, as well as the presence of problematic conditions discussed by various authors globally [16, 18]. Siddiqi [16] portrays urbanization as two-stage process wherein the economic migrants get settled in informal settlements with an informal job first, and then they become permanent resident within the main urban core. By using correlation technique on a data obtained from a random sample, the author explained a positive correlation between education and migration in Pakistan. Lanrewaju [18] assessed the housing quality and impacts of urbanization on environmental degeneration of urban built environment in Nigeria. The author used frequency distribution on secondary data including population census, official documents, and literature to determine urbanization as a root cause of problems such as overburdened infrastructure, overcrowding, and substandard housing, etc.

The new escalated city-making courses and absence of migration profile demand further clarification on this cliché [3, 29]. Scholars are looking scrupulously to the constraints, opportunities, and public reaction with respect to the imposed urban situation worldwide. In the context of third world countries, particularly Pakistan, this study offered new dawn and aims to propose policy recommendations on significant economic and socio-cultural pull-factors via a set of predefined pull-factors derived from the literature [30]. These policy recommendations would make significant contributions in curtailing uncontrolled urbanization and assists in harnessing it before the condition exacerbates. Furthermore, solutions could be replicated in any third world city going through the unfettered urbanization pace and haphazard urban growth.

The upcoming section introduces the study area and its relevance to the topic. It proceeds to reveal the urbanization timeline and discusses why Hyderabad is an eye-catching place for migrants. For a comprehensive background, it continues to describe the tactics and tools exercised to derive the results with accuracy and authenticity. Afterward, the sections of results and discussion highlight the causes of urbanization and its core contributors. At last, the section of policy recommendations grabs the concerns of policy-makers, scholars, and urban planners having a keen interest in calming urbanization pace. It offers optimum suggestions whose inclusion in the policies of urban planning and development, urban management, and urban governance could result in transforming Hyderabad and similar third-world cities to a place that fosters the economy and nurture its urbanites.

2. Methodology

The detailed information regarding the area of interest, the data collection procedure, the data collection instrument, and the data analysis method is discussed hereafter.

2.1. Study Area

The urbanization evidence for this study was obtained from three urban Talukas (a district's subdivision) of district Hyderabad. Hyderabad is Pakistan's 4th largest city and 2nd of the Sindh province, located 24º46' and 26º06' Latitudes and 68º16' and 68º59' Longitudes [31]. As shown in Fig. 1, Hyderabad is divided into 4 Talukas, namely, Hyderabad City, Latifabad, Qasimabad, and Hyderabad Rural. As urbanization is usually observed in the major urban centers, the study was solely confined to urban Talukas excluding suburbs and cantonment areas. The study area can be seen in Fig. 1.

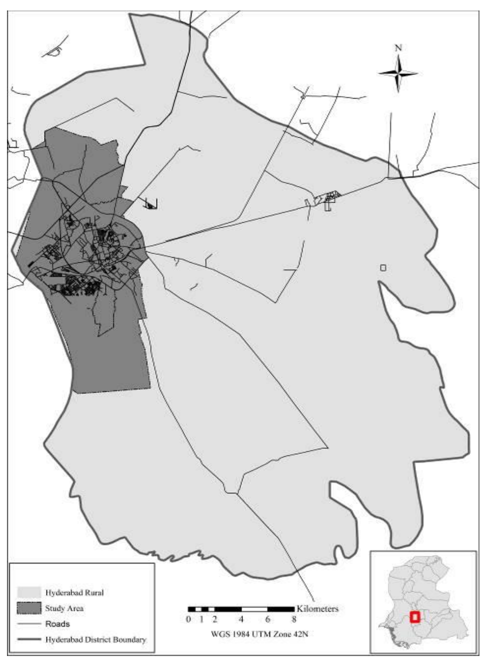


Fig. 1. Map of Hyderabad District

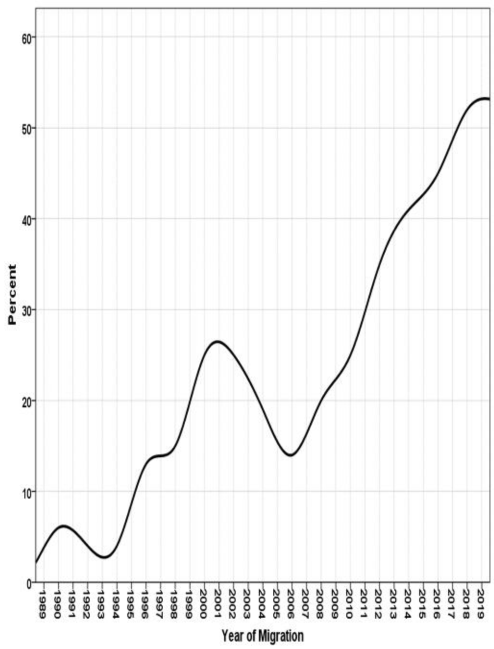


Fig 2. Hyderabad's Urbanization Trend

Since Pakistan's independence, Hyderabad is among those areas that dominate the urbanization scene. Its regional connectivity, urban facilities, and hope for a better life are core reasons of urbanization that contributed 12% of local migrants [32]. Hyderabad's critical setting accelerates urbanization pace ensuing urban glitches stated earlier. In 1998, the district with 8 Talukas had 2,891,488 population and 94,158-lifetime migrants [31]. Whilst, the district with 4 Talukas and 52 Union Councils (the lowest tier of government, a subdivision of Taluka) has 1,834,371 urbanites [27]. Fig.2 shows the recorded years in which the respondents moved to Hyderabad. The graph is significant as it assists to understand the Hyderabad's urbanization trend. As illustrated in Fig. 2, the urbanization pace received a backlash from 2008-2010 due to a severe disturbance in law and order situation. However, the annual flooding from 2010 and onwards along with some better law and order situation gives urbanization a humungous increase and open the doors of immense urbanization towards Hyderabad.

Fig. 2 clearly depicts that the urbanization pace is not stopped yet. The facts are drastically increasing; and poor planning practices, plan implementation's inefficiency, as well as the vacuum created by loopholes and contradictions in governing power, would let it increase on and on [33]. Fig. 3 illustrates a timeline of urbanization-led urban expansion

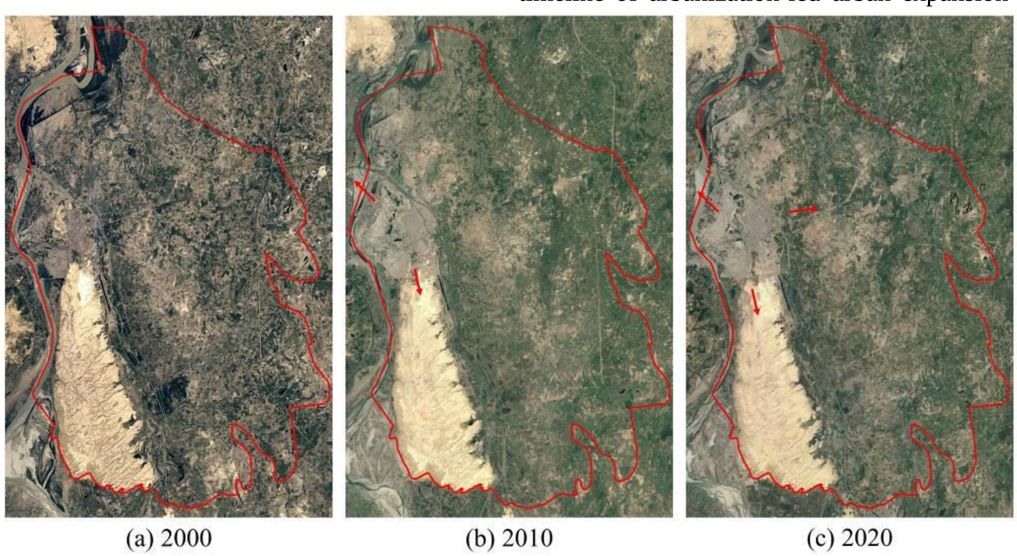


Fig 3. Hyderabad's Urban Expansion Timeline (2000-2020)

in Hyderabad from 2000 to 2020. Fig. 3 (b) clearly shows that in 2010 the city starts expanding towards Taluka Qasimabad and Taluka Latifabad. Whereas, in the year 2020 (Fig. 3 (c)), another dimension (i.e. towards Hyderabad Rural Taluka) was added to the urban expansion. Increase in the density of urban built environment can also be observed in the figure. An important aspect behind this urbanization-led urban expansion is the influences of private real estate developers are complicating the matter. Indeed, they are responding to shelter needs, however, it contributes to the excessive urban population, more dense settlements, and unjustifiable urbanization patterns [34-36]. Like Karachi's urbanization dilemma, poor migrant families usually reside in outlying informal settlements on Taluka boundaries, causing agricultural land misuse and pushing the Taluka boundaries outward [37]. The entire scenario is led by economic migrants and termed as peasant-dominated urbanization, along with the job mobility of impermanent rural migrants [38]. To conclude, economic and socio-cultural pull-factors were found as main drivers for the uncontrolled urbanization process. These clichés convinced the researchers to conduct the study here with an aim to suggest optimum policy recommendations for harnessing urbanization's flow and protecting the area from the urban explosion.

2.2. Theoretical Framework

Following the footprints of previous researches, this study focuses on a multidimensional theoretical framework to explain the economic and socio-cultural aspects of immense urbanization (Gu et al. 2014). The study involves a dependent variable, i.e. urbanization (pull-factors) and a total of 8 independent variables that further categorized into economic, and socio-cultural pull-factors, as shown in Table 1. As poverty is prevailing throughout the country, economic instability is getting acute in rural areas. The rural inhabitants tend to move in search of employment opportunities to central districts. Being an economic hub of Sindh province after Karachi, Hyderabad is the most preferred location for economic migrants to move. Apparently, it offers more employment opportunities either formal or informal in all fields of work. Thus, the probability of getting a job in Hyderabad is more than in other districts of Sindh excluding Karachi. Furthermore, the vast demand for almost every product ensures the continuous supply of the goods and services that makes the job secure. Therefore, economic variables like better employment opportunities (X_1), higher income probability (X_2), and job security (X_3) were selected considering the current economic situation of the country.

In some cases, people migrate to urban areas because they might want to live their life without any socio-cultural taboos, limitations, and restrictions and cities provide shelter to people seeking freedom. Unlike other districts of Sindh province, Hyderabad is preferred as the best place for students due to the presence of numerous excellent schools, colleges, and vocational training institutes; but its proximity to the three best public sector universities of Pakistan could be the most highlighted factor. Likewise, it has better hospitals, better living conditions, better transportation facilities, more amusement places, and easy access to other utilities and services. The easiness in accessing such facilities make Hyderabad a way more comfortable place to live contrary to other districts. Such a scenario results in the selection of socio-cultural variables like desire for independence (X_4), better outlook and hope for the future (X_5), access to basic social services (X_6), comfortable and diverse lifestyle (X_7), and easy access to facilities (X_8). Thus, Table 1 contains a list of predefined variables of urbanization from the literature that emerged as core encouragers of urbanization. Several researchers used bits and pieces of Table 1 to examine the impact of urbanization in their respective areas of interest. However, a collective impact assessment of these variables would be a more comprehensive and multidimensional analysis to understand urbanization from a new perspective. Also, as the major contributors of urbanization with respect to Hyderabad are still unknown, the respective impacts of these variables would help in understanding the urbanization paradigm of Hyderabad. To conclude, pull-factors could be considered as the critical encouragers behind Hyderabad's uncontrolled urbanization. By testing the aforesaid statement, the study intends to propose policy recommendations on significant pull-factors based on the study design discussed henceforth.

Table I. Study Variables

Variables	**Sources**
Better employment opportunities	Turan and Beşirli [2], Akhter [4], Harris and Todaro [6], Zax [8], Crow [11], Ishtiaque and Sofi Ullah [12], Haas and Osland [13], Jones [15], Government of Pakistan [32], Verter and Darkwah [39], George and Shyamsundar [40], McCool and Kruger [41], Amphune and Enaro [42], Cobbinah, et al. [43], Sridhar, et al. [44], Chakma and Akhy [45], Thet [46], Martin and Zuercher [47]
Higher income probability	Harris and Todaro [6], Crow [11], Ishtiaque and Sofi Ullah [12], Lanrewaju [18], Morinière [26], Government of Pakistan [32], Verter and Darkwah [39], George and Shyamsundar [40], McCool and Kruger [41], Sridhar, et al. [44], Thet [46], Grabova, et al. [48], Zweig, et al. [49]
Job security	Crow [11], Lanrewaju [18], Verter and Darkwah [39], Thet [46]
Desire for indepence	Crow [11], Van Noorloos and Kloosterboer [29], Government of Pakistan [32], Thet [46]
Better outlook and hope for the future	Crow [11], Van Noorloos and Kloosterboer [29], Verter and Darkwah [39], McCool and Kruger [41], Amphune and Enaro [42], Cobbinah, et al. [43], Thet [46], Grabova, et al. [48]
Access to basic social services	Turan and Beşirli [2], Akhter [4], Crow [11], Lanrewaju [18], Amphune and Enaro [42], Cobbinah, et al. [43], Thet [46]
Comfortable and diverse lifestyle	Turan and Beşirli [2], Crow [11], Verter and Darkwah [39], George and Shyamsundar [40], Amphune and Enaro [42], Thet [46], Grabova, et al. [48], Zweig, et al. [49]
Easy access to facilities	Turan and Beşirli [2], Crow [11], Jones [15], Amphune and Enaro [42], Chakma and Akhy [45], Thet [46]

2.3. Sampling and Data Collection

The study's sampling plan comprises of mixed research methodology, including, Purposive and snowball sampling techniques for collecting questionnaire survey data [12, 42, 50-56]. The purposive sampling was used to get the response only from the permanently migrated persons, whereas the snowball sampling assists in contacting more migrants as referred by the previous respondent. Due to the unavailability of detailed Census 2017 results, the researchers derived sample size from an estimated present year population. By applying the simplified sample size formula by Yamane [57], 400 migrants' sample at a 95% confidence interval was decided to work on. To collect authentic and detailed information from the respondents, the data were collected using a face-to-face questionnaire survey. The questionnaire was developed under the supervision of academia and field experts. Ethicality was ensured by informing the respondents about the survey purpose. They were also ensured about the anonymization of the questionnaire to void their doubts about privacy. The questionnaire was structured into 2 main sections. Section 1 was intended to collect socio-demographic information, while section 2 recorded their perception regarding the responsibility of pull-factors in causing urbanization on a Likert scale. The 4-point Likert scale based questions were organized to evaluate the migrants' perception from 1=not at all responsible, to 4=completely responsible [58]. Frequency distribution was used to calculate the socio-demographic variables. Furthermore, responses on another variable of urbanization (pull-factors) were recorded on a similar 4-point Likert scale to measure the responsibility of pull-factors in causing uncontrolled urbanization in the perception of migrants. This variable was used as a dependent variable with formerly identified independent variables to construct a multiple regression model to determine the factors influencing urbanization [3, 12, 39, 51, 56, 59, 60]. At last, the collected data was scrutinized to produce results that are discoursed hereafter.

3. Results

Before discussing the impact of urban pull-factors on uncontrolled urbanization, a glimpse on the urbanization profile of Hyderabad, Pakistan, is illustrated in Table 2, for better understanding of results. As illustrated in Table 2, it was found that most of the respondents (54.66%) were young (16-25 years old) when they migrated to Hyderabad, while 39.80% respondents were less than 15 years old, 5.29% respondents migrated in mature age (26-35 years) and only 0.25% migrated in more than 36 years age group. The majority of respondents were university graduates (63.70%), while the rest had attained higher secondary (24.30%), secondary (5.80%), and primary education (4.80%) only. Most of the respondents settled in Qasimabad Taluka, which constituted about 51.64% of the total population. Followed by Qasimabad, 31.39% of respondents settled in Hyderabad city Taluka, while the least (16.96%) settled in Latifabad Taluka. Moreover, most of the respondents had their own houses (54.20%), against rented ones (45.80%) which means that most of the respondents became permanent residents of that area.

Table II. Socio-demographic characteristics of the respodents

Socio-demographic Characteristics		**Percentage**
Age (during migration)	Less than 15	54.66%
	16-25	39.80%
	26-35	5.29%
	More than 36	0.25%
Education Level	University	63.70%
	Higher secondary	24.30%
	Secondary	5.80%
	Primary	6.20%
Migrated Taluka	Qasimabad	51.64%
	Latifabad	16.96%
	Hyderabad City	31.39%
Ownership status	Owned	54.20%
	Rented	45.80%

3.1. Pull-factors Influencing Uncontrolled Urbanization

The measuring rod used for suggesting policy recommendations were the impacts obtained from the statistical modeling technique, i.e. multiple linear regression (MLR). Multiple regression model was estimated to determine the best linear combination of 8 urban pull-factors for predicting their influences in causing uncontrolled urbanization. All economic and socio-cultural pull-factors were found highly responsible for causing uncontrolled urbanization ($p<0.01$), except comfortable and diverse lifestyle that was found significant at $p<0.05$ as shown in Table 3. Model heteroscedasticity was tested using Breusch-Pagan [61] and Koenker [62] tests, where the model was found heteroscedasticity robust.

As the selected economic and socio-cultural pull-factors significantly predicted the dependent variables, the regression model was found a good fit of the data. The R^2 value (0.861) indicates that 86.1% of the variance in the responsibility of pull-factors to cause urbanization was explained by the model. The tolerance value of the coefficients of predictors was recorded well at over 0.139 ($1-R^2$). This shows an extremely low level of multicollinearity among the model predictors. According to the β coefficients represented in Table 3, a significant role of economic and socio-cultural pull-factors was observed in influencing urbanization towards Hyderabad. Where all economic pull-factors (better employment opportunities, higher income probability, and job security) were found statistically significant and emerged as core boosters of urbanization that retain return migration and encourage economic migrants to mingle in the urban fabric. Furthermore, all socio-cultural pull-factors were found statistically significant and appeared as significant propellers of urbanization except desire for independence and access to basic social services as they emerged as negative contributors. Such negative impacts could be justified as the people of Pakistan (especially Sindh) are strongly attached to their cultural ties and solely owns and follows it. They are also free to practice their traditional and religious practices everywhere. Hence people do not seek shelter in cities to escape from their culture and tradition.

Moreover, the basic social services are available in almost every district of Pakistan, especially the neighbouring districts of Hyderabad. Thus, such factors do not influence a person to migrate. By substituting the coefficients of Table 3 in standardized multiple regression equation, the regression model could be expressed as.

$$Y = 0.498 + 0.152X_1 + 0.222X_2 + 0.779X_3 - 0.908X_4 + 0.324X_5 - 0.615X_6 + 0.159X_7 + 0.763X_8 \quad (1)$$

4. Discussion

For many decades, Pakistan has not experienced a dynamic transformation that could push its major cities towards the path of prosperity and development. Indeed, efforts were tempted, but the urban problems are either misunderstood or the role of planning is taken for granted. As a result, people are continuously migrating to urban areas for several reasons. A study in Pakistan revealed that such migration behavior is driven by economic, social, and cultural pull-factors [50]. This situation matches the study findings as many economic and socio-cultural pull-factors were found responsible to frame Hyderabad as a highly dense and urbanized setting. Backing the results of Bahuguna and Belwal [60], Ahmad, et al. [63], Zi [64], Ajaero, et al. [65], the study found economic pull-factors (better employment opportunities, higher income probability, and job security) as

Table III. Regression Coefficients for pull-factors

R = 0.928, R^2 = 0.861, Adjusted R^2 = 0.858, Std. Error of the Estimate = 0.123					
Variables	**Label**	**Unstandardized Coefficients**	**Std. Error**	**t**	**Sig.**
Constant		0.498	0.101	4.930	0.000
Better employment opportunities	X_1	0.152	0.032	4.698	0.000
Higher income probability	X_2	0.222	0.045	4.903	0.000
Job security	X_3	0.779	0.092	8.491	0.000
Desire for independence	X_4	-0.908	0.097	-9.357	0.000
Better outlook and hope for the future	X_5	0.324	0.083	3.919	0.000
Access to basic social services	X_6	-0.615	0.085	-7.236	0.000
Comfortable and diverse lifestyle	X_7	0.159	0.051	3.126	0.002
Easy access to facilities	X_8	0.763	0.052	14.658	0.000

the most common triggers of uncontrolled urbanization because;

i. The secondary and tertiary cities of Sindh are going through the phases of negligence, and urban bias. This gives rise to rural push as people are giving up their hopes and faith from local government that was intended only to facilitate them.

ii. Problems, such as the outdated irrigation system, the water shortage, illegal agricultural water supply, and the monopoly of agriculture market stakeholders are creating hurdles for farmers. As a result, growers are forced to leave their native places and move to urban areas like Hyderabad, as their economic diversity offers many secure employment opportunities. Moreover, as the rural push tends to increase urban pull, which attracts many economic migrants and job seekers towards Hyderabad.

The literature also portrayed socio-cultural pull factors as a prominent reason for rural to urban migration. Studies from Myanmar, Malaysia, and Pakistan revealed hope for better outlook as a strong pull-factor, which resembles the study findings [3, 46, 66]. Unlike Hyderabad, the secondary and tertiary cities of Sindh do not have a variety of urban services. This was clearly observed from the results as easy access to facilities emerged as the strong contributor for immense urbanization. Contrary to Sindh secondary and tertiary cities, Hyderabad possesses many facilities that could enhance the socioeconomic outlook of migrants. It has a better transportation infrastructure. Though public transport is unavailable, still it offers a good communication network within and among its Talukas that increases the interconnectivity of activity centers and promotes industrial productivity. Another study from India found influences of education level and desire of attaining higher education as a strong propeller of urbanization [44]. Results, however, show an agreement with the statement. Numerous quality pedagogy institutes exist in Hyderabad that attracts the youth either to earn by teaching or to study. Also, its proximity to three famous public sector universities, namely, Mehran University of Engineering and Technology (MUET), University of Sindh (UoS), and Liaquat University of Medical and Health Sciences (LUMHS), makes it an optimum location for migrants to reside. In addition, it facilitates migrants from its vast commercial and recreational amenities accompanied by a peaceful atmosphere.

Furthermore, findings indicate the negative impact of desire for independence and access to basic social services. As the joint family system is a common cultural element in almost every community of Pakistan, people often prefer to live with their parents and siblings. Likewise, neither the elders encourage them to live separately, nor they are willing to move. They also own their culture and traditions. Thus, they do not seek shelter in urban areas to live independently. Additionally, as the local governments are ensuring the provision of basic social amenities, people do not prefer to move due to such reasons.

The aforesaid significant pull-factors along with the influences of rural push attract migrants towards Hyderabad. Hope for a better life convinces them to stay as strategic urban locations like Hyderabad are places where dreams come true. As the backlog of preceding pull-factors is rising in Sindh secondary and tertiary cities at an alarming rate. Similarly, urbanization paradigms tend to exceed their tolerable limit, especially in urban settings like Hyderabad, making the situation worse.

To conclude, the analytical findings of the study quantify the economic and socio-cultural impacts on Hyderabad's uncontrolled urbanization. Pull-factors, such as, better employment opportunities (0.152), high income probability (0.222), job security (0.779), better outlook and hope for the future (0.324), comfortable and diverse lifestyle (0.159), and easy access to facilities (0.763), emerged as the significant propellers of urbanization. The emergence of these pull-factors validates the achievement of study's objective. Furthermore, the interpretation of outcomes helped to understand the reasons and consequences of immense urbanization in Hyderabad. Additionally, the study findings deliver a firm baseline to propose optimum policy recommendations. Thus, assists in achieving the study's aim.

5. Policy Recommendations

This study is an initiative towards urbanization control in Hyderabad, which offers a firm foundation to control urban sprawl and haphazard growth. In addition, this research could serve as a helping hand for other urbanization studies conducted worldwide with a special emphasis on third world cities. Considering the principles of feasibility, ethicality, and suitability in proposing policy recommendations, they are structured by keeping in view the major findings and urbanization consequences. The phrase Sindh secondary and tertiary cities refer to Sindh's second-order (Sukkur, and Larkana) and third-order (Mirpurkhas, Nawabshah, Jacobabad, Shikarpur, Tando Adam, and Khairpur) cities. The most optimum policy recommendations that could assist to curtail rapid urbanization and its related issues in the Hyderabad district and other third world urban areas are described henceforth.

5.1. Policy Recommendations for Economic Pull-factors

Policymakers need to understand that facilitating the rural areas with social utilities and services would not halt the migrant's stream except the new openings for the working class are introduced in conjunction with those amenities. The complete implementation of development / strategic / local plans should also be ensured, as the development works can widen employment opportunities for labor and technically-sound people. Also, several practical strategies could be adopted to identify the specialized products of Sindh secondary and tertiary cities along with the establishment of specialized marketplaces for native workers that intensify their potential and reduces the urban-rural income gap that generates 'urban pull'. Advancements in joint ventures between research universities and industrious fiscal divisions could be made to help encourage high-value employment. In addition, provision of more fiscal inducements and professional training in Sindh secondary and tertiary cities could also reverse the urbanization and hold experts and academically qualified personnel in their respective origins. Most importantly, the participation of women and differently-abled persons should be encouraged in income generation activities through the provision of long-term changes in traditional employment values, mainly convenient job natures, like virtual jobs, and remote work, etc. Their expertise should be directed toward information technology (IT), freelancing, innovation, and entrepreneurship to promote businesses and online earning.

Agricultural and agriculture-based industries, like poultry, fisheries, horticulture, dairy should be promoted, and emphasis should give on promoting small-scale mineral, forest-based and agro-processing industrial units and industrial parks. Agriculture based industries may be given special incentives. The agencies like Sui Southern Gas Company (SSGC), Water and Power Development Authority (WAPDA), Water and Sanitation Agency (WASA), and Pakistan Telecommunication Company Limited (PTCL), etc. may direct sanction for the utility services without any delay. Likewise, the non-governmental organizations (NGO's) should be incorporated to establish different cottage and small-scale industries and/or to provide training in migrant generating locations. Similarly, innovative industrial estate patterns in secondary and tertiary cities could be introduced with an industrial tax-free policy to attract investors.

The governing bodies should improve Sindh secondary and tertiary cities as centers of innovation, entrepreneurship, and sources for socioeconomic services and opportunities through prioritizing investments helpful for local economic development, and opportunities for public-private partnerships in urban development and urban services. Also, the local revenue generation and the financial management system of Sindh secondary and tertiary cities should be upgraded to capacitate local economic development. Whereas traditional markets should be modernized to attract more private investors to diversify and revitalize economic growth so that the skilled migrated workforce could be attracted. Similarly, broadening the quantity and extent of high-quality research universities, professional development courses, and internships to incubate youth to earn more in their native places could decrease rural push. Backing up the growth of high-quality research universities can offer more job options and evolving further the entrepreneurial values of migration generating cities. In addition, information regarding accessible fiscal utilities, hoarding programs, and work-related counseling should be exchanged with migrants to boost local commerce in the profitable economic divisions.

5.2. Policy Recommendations for Socio-cultural Pull-factors

The adequate amenities and facilities should be made accessible in Sindh secondary and tertiary cities, while the development authorities should be empowered to manage such facilities to curtail urbanization flow. A most critical element in curtailing urbanization is the formulation of a regulatory framework. Thus, enhancements in rural administrations could be made to effectively deal with issues like rural migration, re-planning, reconstructing, and rehabilitating substandard or insanitary rural communities. Also, improvements in the supply of amenities and services along with a systematic process is the need of time to organize and sustain them. Emphasis should be given to aging and young as their inclusion is crucial to the rural civic vitality and viability.

The decision-makers should evade partiality towards cities. Cities generating a substantial amount of agriculture-based revenue ought to have all essential necessities to support a quality lifestyle. Also, the accessibility and connectivity of remote locations with various primary/metropolitan cities should be ensured. It was found that Hyderabad's civic amenities and facilities, bright city lights, theaters, shopping malls, sports academies, metalled roads, parks, etc. attract thousands of youth from Sindh secondary and tertiary cities. Therefore, the provision of amenities and urban services is much needed. Along with other facilities, lack of educational institutes was also observed as a strong urban pull. Thus, the development of higher education institutions for medical, engineering, and social sciences in Sindh secondary and tertiary cities was proposed to decline migrant's influx. Whereas, maintaining and promoting quality schooling in terms of services and excellent education was proposed to attract young families towards their origins. Also, community-based development was suggested to be empowered to enhance public access to educational opportunities to best harness their indigenous skills. At last, academia and qualified professionals were suggested to increase research practices, especially with the emphasis on Sindh secondary and tertiary cities as a means for finding solutions and new directions for development.

6. Conclusion

The phenomenon of rapid urbanization has become an ever-increasing and alarming urban problem in Pakistan's context as urban development is being prioritized so that the cities could become progressive, resilient, and sustainable. Though urbanization is crucial to attract valuable human resource towards the cities, but it is only fruitful if carried out in a planned and controlled manner. Due to migration-led rapid increase in the urban population of Hyderabad district, almost all vacant land in the urban cores and even open spaces have been metamorphosed into residential use. Now, the trend is affecting the valuable agricultural lands of Hyderabad rural Taluka as the agricultural land is being converted into residential and commercial uses. Consequently, the ever-increasing and unharnessed immense urbanization results in several urban issues of which, overpopulation, sprawl, ever-expanding slums, and over-burdened urban infrastructure, are a few prime examples. To curtail the higher urbanization rate, questions like 'what factors are responsible for this phenomenon?'; 'to what extent they are contributing?'; and 'how urbanization could be controlled?' needs to be answered. This study attempted to answer these questions and aimed to propose policy recommendations via the identification of significant pull-factors by measuring their respective impacts. Contrary to the socio-cultural pull-factors, results suggested that economic pull-factors are more responsible for uncontrolled urbanization in Hyderabad, Pakistan. The economic pull-factors including better employment opportunities, higher income probability, and job security creates a strong urban-pull and portray Hyderabad as an eye-catching destination for economic migrants and job seekers. While socio-cultural pull-factors such as easy access to facilities, better outlook and hope for the future, and comfortable and diverse lifestyle makes urban-pull stronger that draws economically stable families just to enjoy an urban lifestyle.

This study must acknowledge its limitations. Urbanization is a multidimensional process involving 4 factors (pull, push, stay, return) and simply pull-factors standalone cannot unfold the whole urbanization scenario. However, this can be achieved by integrating other factors and their associated variables for better understanding. The study has tried its best to highlight the core encouragers of urbanization in the context of Hyderabad, Pakistan, wherein, the proposed policy recommendations could assist urban planners and policymakers to formulate an urbanization control policy. Furthermore, the methodology could be imitated to estimate the factors of urbanization in other urban areas. Moreover, the statistical model used in this study could be expanded in future to incorporate emerging pull-factors so that the role of urban pull-factors could be understood in several cities of developing countries struggling to cope with this dilemma.

AUTHOR CONTRIBUTION*:* Hassam Bin Waseem (HBW), Mir Aftab Hussain Talpur (MAHT)
Data curation: HBW.
Methodology: HBW, MAHT.
Software: HBW.
Supervision: MAHT.
Writing – original draft: HBW.
Writing – review & editing: HBW, MAHT.

DATA AVAILABILTY STATEMENT: A dataset collected through questionnaire surveys by HBW was used in this study.

CONFLICT OF INTEREST: None

FUNDING: This study was not funded by any public/private funding agency.

Reference

[1] D. Wang, "Socioeconomic characteristics of rural urbanization in Southern Jiangsu, China," *International Journal of Social Economics*, vol. 26 no. 1/2/3, pp. 290-297, 1999.
[2] M. T. Turan and A. Beşirli, "Impacts of urbanization process on mental health," *Anatolian Journal of Psychiatry*, vol. 9, pp. 238-243, 2008.
[3] N. N. A. Aziz, W. H. A. W. Hassan, and N. A. Saud, "The Effects of Urbanization towards Social and Cultural Changes among Malaysian Settlers in the Federal Land Development Schemes (FELDA), Johor Darul Takzim," *Procedia - Social and Behavioral Sciences*, vol. 68, pp. 910-920, 2012/12/19/ 2012.
[4] H. Akhter, "Rural-urban Linkages in Bangladesh: Emerging Scenario of a Secondary Urban Center," *Dhaka University Journal of Earth and Environmental Sciences,* vol. 3, no. 1, 11-08-2014 2014.
[5] A. Shakoori, "Regionalism versus centralism: an analysis of failure of regional planning in Iran," *Journal of Social and Economic Development,* vol. 18, 09/09 2016.
[6] J. R. Harris and M. P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *The American Economic Review,* vol.

60, no. 1, pp. 126-142, 1970.

[7] J. Van Ommeren, G. J. Van Den Berg, and C. Gorter, "Estimating the Marginal Willingness to Pay for Commuting," *Journal of Regional Science,* vol. 40, no. 3, pp. 541-563, 2000/08/01 2002.

[8] J. S. Zax, "When is a move a migration?," *Regional Science and Urban Economics,* vol. 24, no. , pp. 341-360, July 1993 1994.

[9] A. van der Veen and G. Evers, "A simultaneous model for regional labor supply, incorporating labor force participation, commuting and migration," *Socio-Economic Planning Sciences,* vol. 17, no. 5, pp. 239-250, 1983/01/01/ 1983.

[10] B. Swain, "The Impact of Urbanisation on Crop-Livestock Farming System: A Comparative Case Study of India and Bangladesh," *Journal of Social and Economic Development,* vol. 19, pp. 161–180, 2017.

[11] H. Crow, "Factors Influencing Rural Migration Decisions in Scotland: An Analysis of the Evidence," Scottish Government Social Research2010.

[12] A. Ishtiaque and M. Sofi Ullah, "The Influence of Factors of Migration on the Migration Status of Rural-Urban Migrants in Dhaka, Bangladesh," *Human Geographies* vol. 07, no. 2, pp. 45-52, 2013.

[13] A. Haas and L. Osland, "Commuting, Migration, Housing and Labour Markets: Complex Interactions," *Urban Studies,* vol. 51, no. 3, pp. 463-476, 2014/02/01 2014.

[14] A. Kährik, J. Temelová, K. Kadarik, and J. Kubeš, "What attracts people to inner city areas? The cases of two post-socialist cities in Estonia and the Czech Republic," *Urban Studies,* vol. 53, no. 2, pp. 355-372, 2016/02/01 2015.

[15] G. W. Jones, "Population Situation Analysis of Pakistan," United Nations Population Fund (UNFPA)2016.

[16] M. W. Siddiqi, "Rural-urban migration (A case study of Lahore district)," Pakistan research repository. University of the Punjab. Lahore, Pakistan 2004.

[17] S. He, "Consuming urban living in 'villages in the city': Studentification in Guangzhou, China," vol. 52, no. 15, pp. 2849-2873, 2015.

[18] F. Lanrewaju, "Urbanization, housing quality and environmental degeneration in Nigeria," *Journal of Geography and Regional Planning,* vol. 5, no. 16, pp. 422-429, 2012.

[19] C. R. Farrell, "Immigrant suburbanisation and the shifting geographic structure of metropolitan segregation in the United States," *Urban Studies,* vol. 53, no. 1, pp. 57-76, 2016/01/01 2014.

[20] J. Zhu and Y. Guo, "Fragmented Peri-urbanisation Led by Autonomous Village Development under Informal Institution in High-density Regions: The Case of Nanhai, China," *Urban Studies,* vol. 51, no. 6, pp. 1120-1145, 2014/05/01 2013.

[21] D. Webster and L. Muller, "Challenges of peri-urbanization in the Lower Yangtze Region: The case of the Hangzhou-Ningbo Corridor," 2002.

[22] T. G. MacGee, "The Emergence of Desakota Regions in Asia : Expanding A Hypothesis," *The Extended Metropolis : Settlement Transition in Asia,* no. 3-24, 1991 1991.

[23] M. J. Greif, F. N.-A. Dodoo, and A. Jayaraman, "Urbanisation, Poverty and Sexual Behaviour: The Tale of Five African Cities," *Urban Studies,* vol. 48, no. 5, pp. 947-957, 2011/04/01 2010.

[24] M. Qasim, M. M. Anees, U. Ghani, J. Malik, M. Khalid, and A. Bashir, "Environment Degradation Cause By Urbanization in Pakistan (A review Paper)," *Bulletin of Energy Economics (BEE), The Economics and Social Development Organization (TESDO),* vol. 2, no. 3, pp. 62-71, 2014.

[25] D. E. Dowall and P. D. Ellis, "Urban

Land and Housing Markets in the Punjab, Pakistan," vol. 46, no. 11, pp. 2277-2300, 2009.
[26] L. Morinière, "Environmentally Influenced Urbanisation: Footprints Bound for Town?," *Urban Studies*, vol. 49, no. 2, pp. 435-450, 2012/02/01 2011.
[27] (2017). *Provisional Summary Results of 6th Population and Housing Census-2017*.
[28] Government of Sindh. (2008). *Hyderabad Master Plan 2007-27 Pre-Planning Report*.
[29] F. Van Noorloos and M. Kloosterboer, "Africa’s new cities: The contested future of urbanisation," *Urban Studies*, vol. 55, no. 6, pp. 1223-1241, 2018/05/01 2017.
[30] J. R. Logan, "People and plans in urbanising China: Challenging the top-down orthodoxy," *Urban Studies*, vol. 55, no. 7, pp. 1375-1382, 2018/05/01 2018.
[31] (1999). *1998 Census Report of Hyderabad*.
[32] Ministry of Climate Change, Government of Pakistan. (2015). *National Report of Pakistan for HABITAT III*.
[33] Y. Zhan, "The urbanisation of rural migrants and the making of urban villages in contemporary China," *Urban Studies*, vol. 55, no. 7, pp. 1525-1540, 2018/05/01 2017.
[34] R. Grant, "Sustainable African Urban Futures: Stocktaking and Critical Reflection on Proposed Urban Projects," *American Behavioral Scientist*, vol. 59, no. 3, pp. 294-310, 2015/03/01 2014.
[35] D. Hattingh, B. Russo, A. Sun-Basorun, and A. V. Wamelen, "The rise of the African consumer: A report from McKinsey’s Africa Consumer Insights Center," 2012.
[36] KMPG, "The role of cities in Africa’s rise," KMPG International Digital Reports2012.
[37] N. Ahmad, "Choice of Location and Mobility Behaviour of Migrant Households in a Third World City," vol. 29, no. 7, pp. 1147-1157, 1992.
[38] Z. Li and Z. Liang, "Gender and job mobility among rural to urban temporary migrants in the Pearl River Delta in China," *Urban Studies*, vol. 53, no. 16, pp. 3455-3471, 2016/12/01 2016.
[39] N. Verter and S. Darkwah, "Determinants of International Migration: The Nigerian Experience," *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 62, no. 2, pp. 321-327, 2014.
[40] B. P. George and A. Shyamsundar, "The Factors that Affect Indian Migrants' Decision to Stay in or Counter Migrate from the United States: A Study with Special Reference to the Role of Tourism Related Imagery as a Determinant," *Journal of Identity and Migration Studies*, vol. 1, no. 2, 2007.
[41] S. F. McCool and L. E. Kruger, "Human Migration and Natural Resources: Implications for Land Managers and Challenges for Researchers," United States Department of Agriculture, Forest Service, Pacific Northwest Research Station2001.
[42] B. E. Amphune and Y. D. Enaro, "Urban Livelihoods of Rural to Urban Migrants: A Case Study of Wolaita Sodo Town, Southern Ethiopia," presented at the 2nd National Workshop of Research and Community Services, Wolaita Sodo University, 2012.
[43] P. B. Cobbinah, M. O. Erdiaw-Kwasie, and P. Amoateng, "Africa’s urbanisation: Implications for sustainable development," *Cities*, vol. 47, pp. 62-72, 2015.
[44] K. S. Sridhar, R. A. Venugopala, and S. Pavan, "Is it Push or Pull? Recent Evidence from Migration into Bangalore, India," *Journal of International Migration and Integration*, vol. 14, no. 2, pp. 287–306, 2013.
[45] B. M. Chakma and A. A. Akhy, "Hill to

Plain: Causes and Impacts of Internal Migration of Indigenous Peoples in Bangladesh," *Sociology Mind,* vol. 05, no. 04, pp. 268-275, 2015.

[46] K. Thet, "Pull and Push Factors of Migration A Case Study in the Urban Area of Monywa Township , Myanmar.pdf>," *World of Statistics,* vol. 1, no. 4, 2014.

[47] P. Martin and G. Zuercher, "Managing migration: The global challenge," *Population bulletin,* vol. 63, pp. 3-20, 03/01 2008.

[48] P. Grabova, B. Dosti, and E. Pojani, "Effects of Economic Crisis on Migration - Evidence from Albania," *European Scientific Journal,* vol. 9, no. 22, 2013.

[49] D. Zweig, C. S. Fung, and D. Han, "Redefining the Brain Drain," *Science, Technology and Society,* vol. 13, no. 1, pp. 1-33, 2008.

[50] N. Farah, M. I. Zafar, and N. Nawaz, "Socio-Economic and Cultural Factors Affecting Migration Behavior in District Faisalabad, Pakistan," *Pakistan Journal of Life and Social Sciences,* vol. 10, no. 1, pp. 28-32, 2012.

[51] J. G. Adewale, "Socio-Economic Factors Associated with Urban-Rural Migration in Nigeria: A Case Study of Oyo State, Nigeria," *Journal of Human Ecology*, vol. 17, no. 1, pp. 13-16, 2005/01/01 2005.

[52] E. Kerselaers, E. Rogge, E. Vanempten, L. Lauwers, and G. Van Huylenbroeck, "Changing land use in the countryside: Stakeholders' perception of the ongoing rural planning processes in Flanders," *Land Use Policy,* vol. 32, pp. 197-206, 2013.

[53] G. Thondhlana, "Land acquisition for and local livelihood implications of biofuel development in Zimbabwe," *Land Use Policy,* vol. 49, pp. 11-19, 2015.

[54] T. A. Tran, "Land use change driven out-migration: Evidence from three flood-prone communities in the Vietnamese Mekong Delta," *Land Use Policy,* vol. 88, 2019.

[55] Abbas Tashakkori and C. Teddlie, Abbas Tashakkori and C. Teddlie, Eds. *Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications, 2013.

[56] A. Joshua Remi, A.-A. Ibraheem Adegoke, and O. Nurain. A, "An Appraisal of the Factors Influencing Rural-Urban Migration in Some Selected Local Government Areas of Lagos State Nigeria," *Journal of Sustainable Development,* vol. 4, no. 3, 2011.

[57] T. Yamane, *Statistics : an introductory analysis* (no. Accessed from https://nla.gov.au/nla.cat-vn1136113). New York: Harper and Row, 1967, p. 915.

[58] J. Hartley, "Some thoughts on Likert-type scales," *International Journal of Clinical and Health Psychology,* vol. 14, no. 1, pp. 83-86, 2014.

[59] G. K. Uyanık and N. Güler, "A Study on Multiple Linear Regression Analysis," *Procedia - Social and Behavioral Sciences,* vol. 106, pp. 234-240, 2013/12/10/ 2013.

[60] P. Bahuguna and O. K. Belwal, "Regression Model Approach for Out-Migration on Demographic Aspects of Rural Areas of Pauri Garhwal," *International Journal of Management and Social Sciences Research (IJMSSR),* vol. 2, pp. 175-182, 2013.

[61] T. Breusch and A. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica,* vol. 47, no. 5, pp. 1287-94, 1979.

[62] Roger Koenker and G. B. Jr., "Robust Tests for Heteroscedasticity Based on Regression Quantiles " *Econometrica,* vol. 50, no. 1 pp. 43-61, 1982.

[63] N. Ahmad, Z. Hussain, M. H. Sial, I. Hussain, and W. Akram, "Macroeconomic Determinants of

International Migration from Pakistan " *Pakistan Economic and Social Review*, vol. 46, no. 2, pp. 85-99, 2008.

[64] L. Zi, "Urbanization and Economic Growth in China—An Empirical Research Based on VAR Model," *International Journal of Economics and Finance,* vol. 9, p. 210, 02/20 2017.

[65] C. Ajaero, I. Madu, and A. Mozie, "Appraisal of The Factors of Rural-Urban Migration In Southeastern Nigeria," *Innovare Journal of Social Sciences,* vol. 1, pp. 1-8, 12/21 2013.

[66] N. Ahmad, Z. Hussain, M. Sial, H. Ijaz, and W. Akram, "Macroeconomic Determinants of International Migration From Pakistan," *Pakistan Economic and Social Review,* vol. 46, pp. 85-99, 04/01 2008.

Vol. 5, No. 1 | January – June 2021

Design of Isolated Micro Grid Using Renewable Energy Resources with Energy Storage System via HOMER Software A Case Study of a Thar Community

Mahesh Kumar[1], Waleed Memon[1], Amir Mahmood Soomro[1]

Abstract:

Electricity is major requirement for each individual or group to live a prosper and good life, while unfortunately in Pakistan still there are some areas which does not have provision towards this facility. The microgrid is designed to power up those rural areas of Pakistan which does not have provision of electricity via conventional means. To achieve this feat this paper presents a case study of a Thar community or small village located in Thar, Sindh, Pakistan. The small village comprises of 40 houses, some shops and a primary school. The study area having latitude and longitude of 240 52.7' N and 700 14.4' E respectively. The microgrid consists of a Diesel Generator, Photo voltaic solar panels, Batteries and Flywheel systems. The purpose is to find the most optimal solution using different parameters via Homer optimization software.

Keywords: *Optimization, Photovoltaic, Homer, Energy, Optimization*

1. Introduction:

According to World Energy Outlook study by International Energy Association [1], there are 51 million people in Pakistan who does not have access to electricity, from which 63% people are from rural areas. The cause of this situation is due to energy crisis and huge gap between supply and demand in Pakistan. The reason to quote this energy state problem is that we had relied from past two decades to non-renewable energy resources and till now we are relying on it. While other nations have changed their course from non-renewable energy resources to renewable energy resources for the betterment of environment, to conserve non-renewable energy resources and to tackle the ever increasing demand of energy. In order to achieve the foresaid statement an isolated microgrid is to be designed using PV solar panel and diesel generator with and without energy storage systems to obtain a feasible and optimized solution for the community using homer software by aiming the highest penetration of renewable energy resource in the system.

A research paper was presented on standalone residential PV solar system for Assiut city. In this he proposed an isolated PV panel's based optimized and feasible solution containing an estimated load of 2 kW (lighting load, washing machine, TV and other small appliances load) and proposed the most feasible and optimized solution [2]. Another isolated microgrid for small community of cholistan using PV panels and bio mass gas as renewable energy resources and calculated the feasibility report via homer software [3].

A research paper was presented consisting a hybrid microgrid system using wind, solar and hydrogen renewable energy resource to meet the load in most feasible and optimized way. They proposed isolated as well as on grid provision in the system, while all cost energy optimization reports

[1]Department of Electrical Engineering, Mehran University of Engineering and Technology Jamshoro, 76062, Sindh Pakistan.

Corresponding author: rathii.mahesh@faculty.muet.edu.pk

were established using homer software [4]. Another research paper as proposed by an Indian author he presented a decentralized micro grid system in India, Kondapalli forest on Indrakeeladri hillocks of Vijayawada using PV panels and wind turbine, for this system the optimized values and feasibility is obtained using homer software [5].

An optimal hybrid renewable system for isolated systems i.e. three remote islands located in Maldives using three different resources PV Panel, Wind Energy and Diesel Generators. The optimization and feasibility levels are achieved by using Particle Swarm Optimization (PSO) technique to calculate the PV panels, Wind Energy and Diesel Generator parameters, Carbon emissions, fuel severity analysis, and solar irradiation for that area [6]. A research paper was proposed by an Iranian author he proposed a hybrid microgrid system for a village located in Iran, currently having diesel generators to meet the load demand. He introduced to incorporate wind and solar renewable energy sources in conjunction with existing diesel generators to obtain a feasible and optimized solution. The system parameters will be defined for standalone as well as on grid configuration to effectively utilize the resources [7].

A distributed microgrid for small community whom there were no conventional power feasibility was available. The proposed microgrid was based on PV arrays, wind energy, small hydro turbine along with battery backup systems, all the feasibility and optimization calculations were performed using homer pro software [8]. Another hybrid microgrid concept using micro hydroelectric (MHP) source and PV arrays to tackle the ever increasing demand of electricity in Indonesia via renewable energy resources. In order to achieve this feat they used homer and matlab software for simulation and calculation to obtain the most feasible and optimized solution [9].A research paper was concluded using a standalone hybrid microgrid system for a university of Mombasa consisting of small appliance load to heavy machinery loads by using PV arrays and wind turbine renewable energy resources. The feasibility and optimization calculations are achieved using homer software [10-14].

2. Research Methodology

2.1 Community Load Profile

The village consists of 40 houses, few shops and primary school. Unlike urban areas these houses are distant from one another and mostly comprises lighting, fans and domestics electric motors for water pumping. The average estimated load for the community is 535.2 kwh/d and a peak demand of 61.47 kW. The daily load profile graph was shown in Fig. 1 and monthly average load demand graph is shown in Fig. 2.

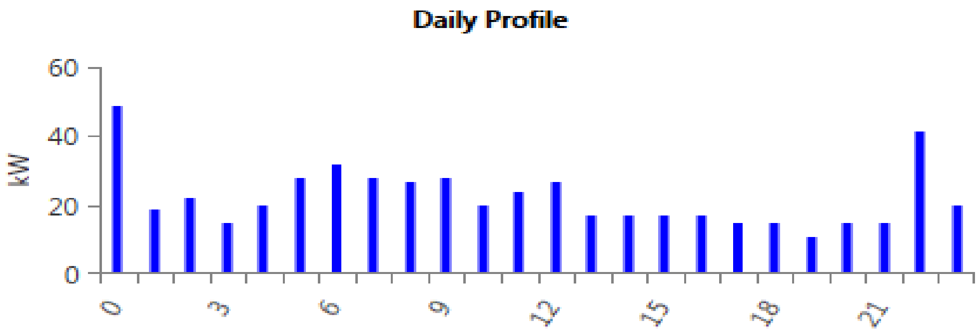


Fig 1:Daily Load Profile of Community

2.2 Solar PV Module

The solar irradiance for the described community was calculating using HOMER database system, the location coordinates were 240 52.7' N latitude and 700 14.4' E longitude. The scaled annual average and clearance index was calculated by HOMER software. The scaled annual average for the community load was found to be 5.61 kwh/m2 per day while average clearance

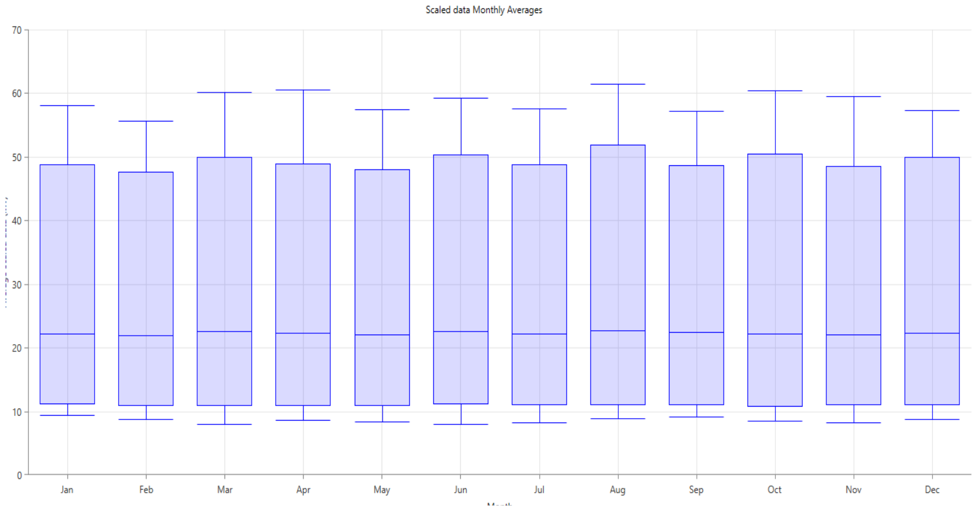


Fig. 2 Monthly Average Load Data of the Village/Community

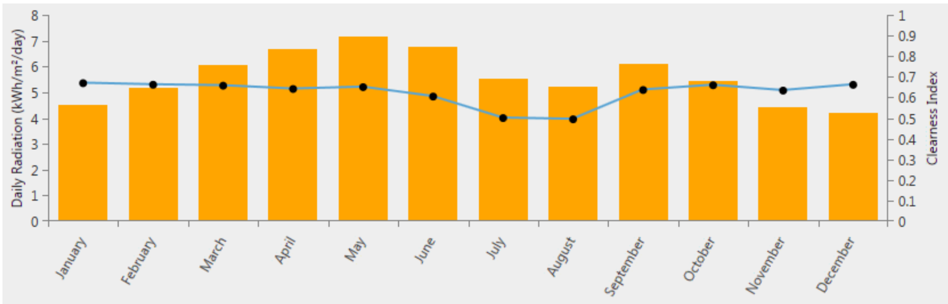


Fig. 3 Solar Irradiance Graph for the Defined Community

index in 0.62. Fig.3 represents the solar irradiance data for the defined community.

2.3 Battery

Batteries are used to store electricity generated by renewable energy resource in order to utilize that electricity when renewable energy is not available in case of solar during night time. This to ensure availability of power. The type of battery used here is 83 Ah, 1 kwh, VRLA batteries of 12 volts, with life expectancy of 10 years and operation and maintenance cost of 10 US$/year. The capital cost of battery is 300 US$ and Replacement cost is 225 US$.

2.4 Converter

A converter is an electrical device which transforms ac power into dc via rectifier and dc power tp ac via inverter, the converter is used here is a bidirectional inverter/rectifier. The cost of this bidirectional convertor is 300 US$/kW of power conversion with replacement cost of 225 US$/kW. The convertor having 95 % rectification and inversion efficiency while the cost of O&M is around 15 US$/year.

3. Simulation Final Results and Discussions

The optimization results of microgrid is shown in Fig. 4, the optimization results are acquired using HOMER software. There are two systems compared and analyzed on COE and total NPC. The systems are compared with and without energy storage systems and the results were analyzed and most optimizes solution is acquired in terms of COE and total NPC with the aim of highest renewable energy penetration in the system. The first system comprises only PV modules, Diesel Generator and Convertor as shown in Fig.5. The lowest NPC for the system is 944,156 US$ while cost of electricity per kWh is 0.374 US$.

The optimization result for the system with energy storage system is obtained and shown in Fig.6, this configuration contains PV module. Diesel Generator, Convertor and ESS (Batteries and Flywheel system). The most optimized solution obtained from this configuration is with total net present cost of 644,201 US$ and cost electricity per kWh of 0.255. This system is much feasible in terms of total NPC and COE than the system without ESS, secondly this system has highest renewable energy resource penetration which can be seen in Fig. 7

Sensitivity	Architecture								Cost				System		Gen Set					
Diesel Fuel Price (US$/L)					PV (kW)	Gen Set (kW)	Converter (kW)	Dispatch	COE (US$)	NPC (US$)	Operating cost (US$/yr)	Initial capital (US$)	Ren Frac (%)	Total Fuel (L/yr)	Hours	Production (kWh)	Fuel (L)	O&M Cost (US$/yr)	Fuel Cost (US$/yr)
0.750					20.0	60.0	10.0	CC	US$0.374	US$944,157	US$69,399	US$47,000	7.25	60,556	8,760	181,181	60,556	10,512	45,417
1.25					30.0	60.0	20.0	CC	US$0.528	US$1.33M	US$98,594	US$60,000	9.21	59,586	8,760	177,347	59,586	10,512	74,482
1.50					30.0	60.0	20.0	CC	US$0.605	US$1.53M	US$113,491	US$60,000	9.21	59,586	8,760	177,347	59,586	10,512	89,378

Fig. 4 Optimal solution Table

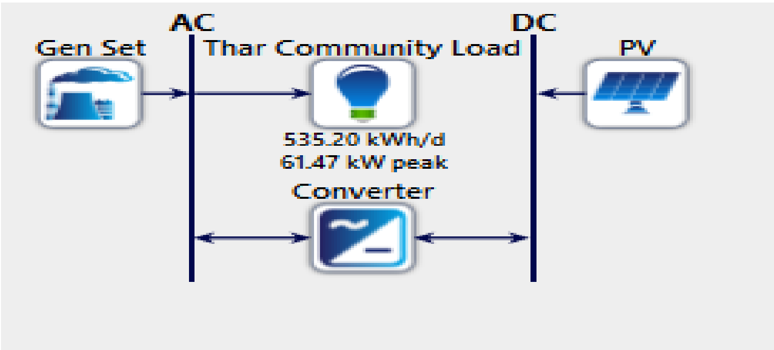


Fig. 5. Systematic Diagram of system without ESS

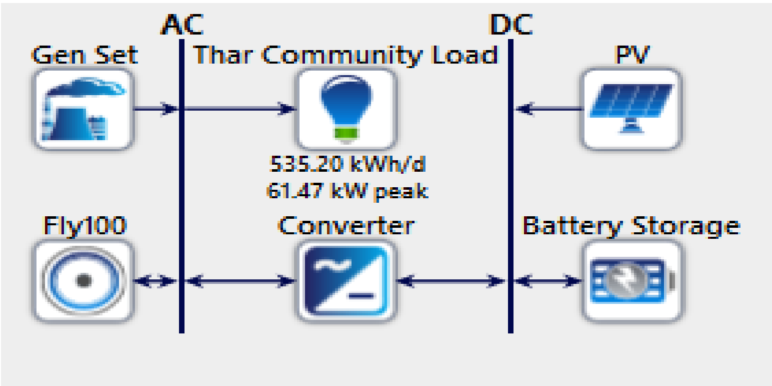


Fig.6. Systematic Diagram with ESS

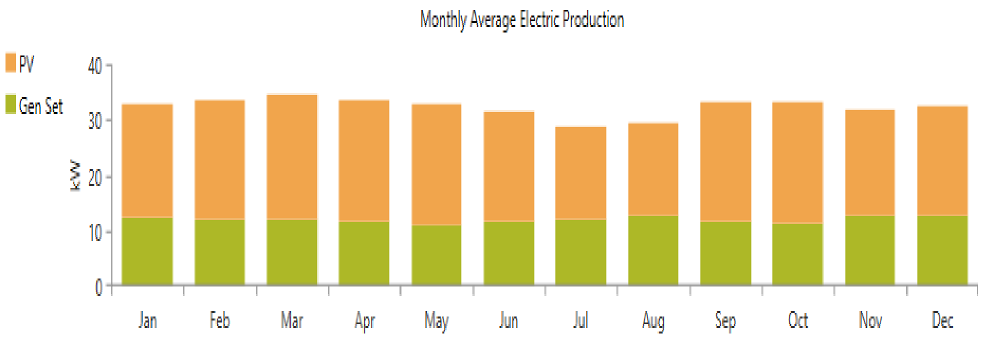


Fig. 7. Electricity production share by PV and Diesel Generator

4. Conclusion

The results obtained by HOMER software clearly depicts that the system with energy storage system ESS having lowest total net present cost and lowest cost of energy COE per kWh generated. The system also comprises highest amount of renewable energy resource penetration that of 63 % percent of electricity is generated using PV solar panels while 37 % of electricity is generated via diesel generator. The generation share in terms of kWh/year for PV panels and by diesel generator was 178,004 kWh/year and 104,427 kWh/year making a total generation of 282,431 kWh/year. Thus contributing less carbon emission in the environment with lowest possible energy cost. This also fulfils the need to electrify the rural areas where electricity provision in not possible by national grid via hybrid energy resources.

Acknowledgment

The authors thanks Mehran University of Engineering and Technology Jamshoro for providing research facilities.

References

[1] Vincietal, S., "Energy Access Outlook 2017: From Poverty to Prosperity," International Energy Agency, Volume 94, No. 3, pp. 116, 2017.

[2] Abdel-Salam, M., Ahmed, A., Amery, M., Swify, M., El-kousy, A. and Sayed, K., 2011, December. Design and implementation of stand-alone residential PV system. In 2011 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1-6). IEEE.

[3] Türkay, B.E. and Telli, A.Y., 2011. Economic analysis of standalone and grid connected hybrid energy systems. Renewable energy, 36(7), pp.1931-1943.

[4] Krishna, K.M., 2011, December. Optimization analysis of microgrid using HOMER—a case study. In 2011 Annual IEEE India Conference (pp. 1-5). IEEE.

[5] Oulis Rousis, A., Tzelepis, D., Konstantelos, I., Booth, C. and Strbac, G., 2018. Design of a hybrid AC/DC microgrid using HOMER Pro: case study on an islanded residential application. Inventions, 3(3), p.55.

[6] Aly, A.M., Kassem, A.M., Sayed, K. and Aboelhassan, I., 2019, February. Design of Microgrid with Flywheel Energy Storage System Using HOMER Software for Case Study. In 2019 International Conference on Innovative

Trends in Computer Engineering (ITCE) (pp. 485-491). IEEE.

[7] Asrari, A., Ghasemi, A. and Javidi, M.H., 2012. Economic evaluation of hybrid renewable energy systems for rural electrification in Iran—A case study. Renewable and Sustainable Energy Reviews, 16(5), pp.3123-3130.

[8] Hafez, O. and Bhattacharya, K., 2012. Optimal planning and design of a renewable energy based supply system for microgrids. Renewable Energy, 45, pp.7-15.

[9] Nazir, R., Laksono, H.D., Waldi, E.P., Ekaputra, E. and Coveria, P., 2014. Renewable energy sources optimization: A micro-grid model design. Energy Procedia, 52, pp.316-327.

[10] Sandeep, G. and Vakula, V.S., 2016, March. Optimal combination and sizing of a standalone hybrid power system using HOMER. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 4141-4144). IEEE.

[11] Gilal, A. R., Jaafar, J., Omar, M., Basri, S., & Waqas, A. (2016). A rule-based model for software development team composition: Team leader role with personality types and gender classification. Information and Software Technology, 74, 105-113.

[12] Gilal, A. R., Jaafar, J., Capretz, L. F., Omar, M., Basri, S., & Aziz, I. A. (2018). Finding an effective classification technique to develop a software team composition model. Journal of Software: Evolution and Process, 30(1), e1920.

[13] Kumar, M., Das, B., Baloch, M. H., Nallagownden, P., Elamvazuthi, I., & Ali, A. (2019). Optimal Placement and Sizing of Distributed Generators and Distributed-Static Compensator in Radial Distribution System: Distributed Generators and Distributed-Static Compensator. International Journal of Energy Optimization and Engineering (IJEOE), 8(1), 47-66.

[14] Kumar, M., Nallagownden, P., & Elamvazuthi, I. (2017). Optimal placement and sizing of renewable distributed generations and capacitor banks into radial distribution systems. Energies, 10(6),811.

Vol. 5, No. 1 | January – June 2021

Challenges Faced in the Collection and Disposal of Municipal Solid Waste: A Case Study of Sanghar City

Laraib Bhatti[1], Mir Aftab Hussain Talpur[1], Irfan Ahmed Memon[1], Imtiaz Ahmed Chandio[1], Fahad Ahmed Shaikh[1]

Abstract:

The generation of municipal solid waste in many cities has increased as the population is increasing. Mismanagement of MSW collection is creating, health hazards, socio-economic problems, and deteriorating environments. This research aimed to examine the SWM problems and to identify SWM issues through the inhabitant's perspective and propose a sustainable method to control, collect, treat, utilize and disposing of Municipal SW in Sanghar City. Hence, the objective of this study is to achieve through cluster sampling, for the residential questionnaire study, a sample size of 384 was obtained and descriptive statistics and regression and correlation analysis methods have been used to evaluate the data collected. As a result, the collection of MSW is quite inappropriate and constricted to the influential area and solid waste keeps on scattered throughout the remaining areas. In recent years, the mismanagement of MSW has become a major problem in Sanghar city. The key problems of SW in the city are indiscriminate disposal, improper collection, inadequate storage, and insufficient facilities. To address these issues, the management of the disposal of SW must be carried out with the complete participation of the respective communities.

Keywords: *Municipal solid waste; collection; disposal; quality of life; residential satisfaction.*

1. Introduction

In general, MSW is a collection of commercial and household waste created by the living population [1]. Municipal Solid Waste contains recycled paper, cans and bottles, food waste, yard trimmings, and other products [2]. Municipal solid waste comprises any human-derived waste that is usually unwanted as unnecessary or discarded [3]. MSW is a substance that is no longer useful to the individual responsible for it, nor is it meant to be released directly in a pipe. Normally this does not contain human excreta [4]. MSW is produced domestically and commercially in public places and streets. The words “garbage”, “trash”, “refuse” and “rubbish” were used to refer to some categories of MSW. MSW is therefore any raw material that comes from residential and commercial operations and is deemed to be discarded by those who own it. It takes us to municipal solid waste sources [5].

MSW is one of the big environmental issues facing cities in many underdeveloped countries, including Pakistan. The growth of the economic and urban population is leading to an increased generation of municipal SW. The use of products that cause toxic waste is

[1] Mehran University of Engineering and Technology, Jamshoro Sindh Pakistan
Corresponding Author: blaraib01@gmail.com

another burden. Unmanaged disposal of city solid waste adds to pollution and general prosperity risks in the districts. Especially in low-income groups, most regions lack a collection of refuse. Locals in low-income societies tend to dump or burn their trash in close empty areas, particularly government sites and rivers [6]. The highest generation of waste of the Kumasi metropolitan area was recorded at 0.75 kg/person/day [7]. Due to the absence of laws and regulations, fundraising, self-awareness, management and knowledge, and the collection of cars, machinery, and recycling technology, it is unable to manage the continually increasing waste products quantity [8].

Therefore, Proper management of MSW is important for minimizing the impacts of environmental conditions and land degradation. Proper management of MSW has become a big issue among many cities in the developing world, and Pakistan is one of them [9]. It is assumed that if MSW is properly handled, it can be a beneficial resource, but if it is not managed efficiently, it can become a cause of environmental and human hazards. More to the point it is also believed by different institutions that is one of the most important components parts of urban sanitation is municipal SW management.
There is a lack of planning and management in Sanghar City. There are no appropriate arrangements for the management of SW. This research concentrates on the present SW management framework especially in the district and its adverse effects on condition and to survey the system, practices, and obligations of different organizations required in strong waste administration inside the premises of the city zone.

2. Literature Review

Improper management of SW creates different environmental and health problems for local people in cities. Mechanisms of municipal SW management in developed and underdeveloped counties are also discussed in this review. This literature may use to accomplish research objectives and give support to suggest a better mechanism of municipal solid waste management to solve overcome problems in the study area. The rapid population growth and urbanization process translate into more waste created [10].

Another research in Kolkata, India, found that a lack of facilities exists, and those improper bin collections are to blame for inadequate municipal solid waste collection and transportation. These studies have shown that this issue must be addressed because it directly affects the climate and culture.

2.1. Municipal Solid Waste

It is a global challenge to manage environmentally sustainable municipal solid waste [11]. However, many municipalities fail to control solid waste management issues, due to a not properly managed system [12]. While some administrated bodies have formed environmental protection plans of action, rules, which are, unfortunately, just take on only in capital cities [6]. Open space areas are used for dumping in most cities [12].
Reducing this problem is the biggest challenge for developing nations such as Pakistan. The public playing a significant part in reducing and managing this problem in general. In Pakistan, it is revealed that there are nearly produced 64,000 tons per day of solid waste [13]. These come out to be an insufficient understanding of the problem of solid waste management (SWM), whereas the first and efficient management of waste is just feasible to a public behavior [14]. Japan is reported to be effective in decreasing solid waste by applying a shared accountability notion where the public separates the waste before dumping it into fuels, non-burnable, and recyclables [15].

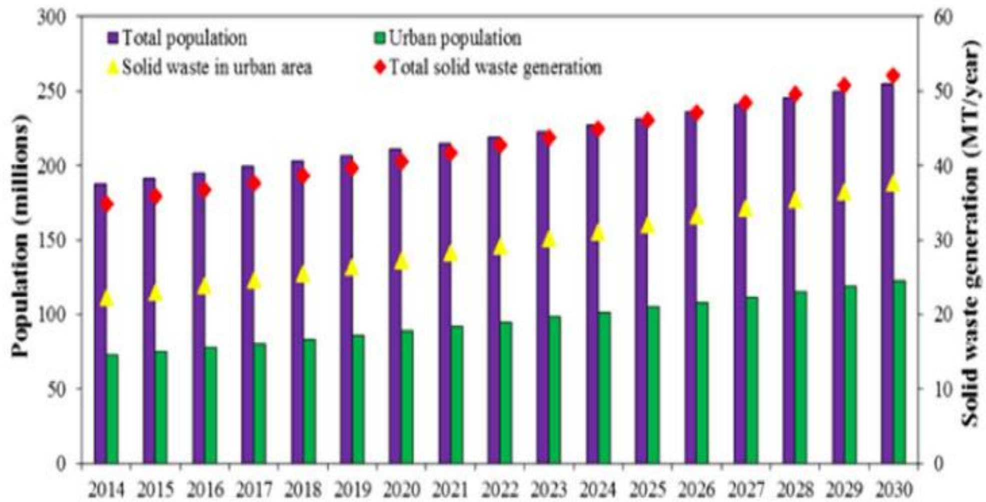


Fig 1: Estimation of Population and generation of MSW in Pakistan (Kawai and Tasaki, 2016)

2.2. Comparison of the activities of MSW management in Pakistan and other countries.

Those countries with a lower GDP produce less SW [16]. By refusing this statement, it explains that SW generation is gathered from information provided by a variety of sources in cities, considering municipalities, NGOs, research centers, higher education institutions, or even the first author. Whereas GDP is a national economic measure. Table 1 displays the generation of MSW coming from various regions and countries.

The amount of MSW produced increases as the population increase, as shown in Figure 1. MSW generation per capita is inversely proportional to national economic growth and varies by country.

3. Material and Methods

The data was gathered from various primary and secondary sources of data. Secondary data were gathered using comprehensive sources, i.e. literature studies in books and journals written. I have selected district Sanghar City and the data have been obtained and analyzed in SPSS 22.0 using descriptive analysis and the technique of regression and correlation analysis. The data identified the main critical causes for the city of Sanghar and discussed their possible solution with the field expert.

The area selected for research purposes is Sanghar city to examine significant issues of MSW management. Sanghar is the headquarters of Sanghar district and Sanghar Taluka. As per the census, 2017 total population of Sanghar city is four, 34,087. The total area of Sanghar city is 2,218 square kilometers. The Coordinates of Sanghar city is 26°2'49" N 68°56'54"E. Generating Solid waste is 350gm per person per day. The total generation is over 250 tons. Sanghar has located approximately 265km from Karachi.

Table I: Generation rate of MSW in different countries

Study year	Region/Country	GR (kg/capita /day)
(2015)	High Income	2.1
	Upper Middle	1.2
	Lower Middle	0.79
	Lower Income	0.6
	African	0.74
	Arabs States	0.77
(2011), (2015)	Asia	0.79
	Latin America	0.82
	Industrial Nations	1.4
	Transition Nations	1.34
	All cities Average	0.96
(2011)	Beijing (China)	1.2
(2010)	Singapore (Singapore)	0.96
(2012)	Dutse (Nigeria)	0.97
	Katsina (Nigeria)	1.12
	Denmark	2.04
	France	1.45
	Netherlands	1.44
(2016)	Kuwait	1.4
	United Kingdom	1.32
	Sweden	1.25
	Romania	0.74
(2012), (2015)	Pakistan	0.57
	Mumbai	0.45
	Kolkata	0.58

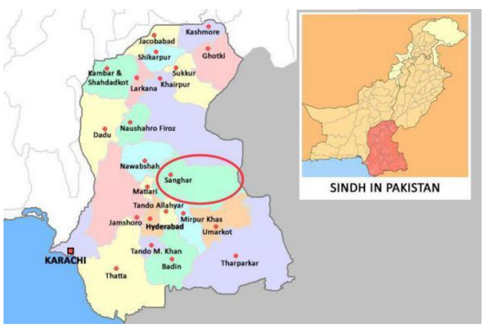


Fig 2: Sanghar district Map (Source: Google Map)

3.1. Method

The data was collected from a field survey through various techniques i.e. quantitative methods, a detailed personal field survey. Besides, the quantitative method was utilized to obtain detailed information about "municipal solid waste management" and its impacts on research area residents. A uniform questionnaire was developed and collected from the selected area residents. Detailed survey for municipal solid waste carried out through personal field surveys, personal interviews, questionnaires. The questionnaire may contain open-ended or close-ended questions depending on the nature of the research.

During the day, the survey was conducted for the residents at their homes. The aims of the survey were mentioned to potential residents during the survey [17]. The purpose of this data collection was to find out people's views on the impacts of inadequate SWM on the environment and human health. This strategy has helped test public perceptions of the issue of SWM. It also helped to determine whether their perceptions of the impact that the improper management of SW can have on the environment, likewise on human health is reliable.

Many unnecessary materials are produced in every area of human life, and these materials are discarded simply because they are considered waste. Waste is a core problem in developing countries where waste generation per production unit is far higher than in developed countries due to unsustainable manufacturing practices, poor design, and bad decision-making eventually. The goal of this study was to identify the problems with SWM in Sanghar and its effect on SW management practices. This work is investigative, as well as casual with a total population of 434087 a sample size of 384 is used. The self-administered Questionnaire instrument has been used to collect study data.

The analyzes of the collected data are based on quantitative data analysis techniques (such as mean, percentages, ratios, and standard deviation) and qualitative data analysis techniques (such as content analysis). It was found that not all of the issues facing SWM activities were considered major challenges but the only institutional structure and appropriate laws on SW management. The impact of the problems on waste management practices on the results leads to inefficiency in solid waste management practices. Based on this study's results, it is suggested that management adopt steps to address the problems facing SWM practices.

3.2. Sample Population

The standard sample population and size of the households must be calculated. The standard level of the total population is (5-10%) which can be used as a sample of the total population [18].

TABLE II. Population Sample (Estimated)

Taluka	Population in 2017	Total Population 2021 (projected)	5% Sample Standard Selecti	Number Of Sampled Households (Questionnaires)
Sanghar	4,34,087	4,39,319	384	384

Sanghar population was 434087 as per Census 2017 with 50% population proportion and the questionnaire sample size is 384 which is taken in (Krejcie and Morgan 1970) table [19]. 5% sample standard selection is taken in the(Israel 1992) table [20]. Clusters from the study area were chosen, which have at least 10 residential units and possessed similar socioeconomic characteristics. As one questionnaire would be able to represent 10 households and approximately 60 inhabitants. Similarly, 384 questionnaires would exemplify 3840 households and 23040 persons in the study area. With the help of quota sampling, this proportion did take into account to satisfy the standard levels of population and sampled households. Thus, the proportions were made and cluster-sampling methodologies were implemented [21-29].

To conduct data about the existing mechanism of municipal solid waste management, machinery, workforce, existing duping points, mod of municipal solid waste collection, Tehsil Municipal Administrations (TMA's) Rules, and Bylaws the interviews were conducted from different officials of concerned authority working on MSW management. To also conduct interviews to know about a previous study on municipal SW management in the city. The following detailed information was received from the authorities concerned about their department [30-34].

4. Results and discussion

Two waste disposal sites have been chosen and visited to observe the problems surrounding municipal solid waste management (SWM) in Sanghar, as well as to understand the sources and forms of waste disposal.

2.3. Sources of Municipal Solid Waste

Municipal SW sources are categorized into two categories of residential and commercial waste. This categorization was based on observations made during visits to the site. Each one source led to various forms of residues. Waste forms found during disposal station visits included cardboard, batteries, paper, wood, plastics, glasses, etc. The details of the waste found at the disposal site concerning its generation source are discussed in Table III.

TABLE III. Sources and Types of Municipal Solid Waste

Source	Typical Location	Types of Solid Waste
Residential	House/ Apartments	Food waste, cardboard, chemicals, bottles, metals, textiles, dust, paper, special waste (bulk foods, consumer electronics

		, batteries, oil, and tires), and household hazardous waste.
Commercial/Municipal	Offices, restaurants, hotels, markets, and stores	Paper, cardboard, iron, wood, plastics, food waste, glass, Special waste, toxic waste

Structured questionnaires and unstructured interviews were conducted at the disposal sites during the visit with the staff and the individual responsible for managing the waste. By analyzing 384 questionnaire forms obtained during these visits, the respondents were statistically analyzed to reach a consensus. The objective of the questionnaire and the interviews were conducted to find explanations for the production of waste that was disposed of at the selected sites. The results obtained from the Likert questionnaire scale are discussed in Table IV.

The mean of the first factor has 0.972 SD (Standard Deviation). Most of the respondents i.e. 281 out of 384marked this factor strongly agree. The major reason is, an increase in population is a common cause of waste generation in Sanghar city and most people agree with it that is why it has the highest value of RII too. For poor management, the factor has 0.945 SD (Standard Deviation). Most of the respondents i.e. 103 out of 384 marked this factor strongly agree. For factors second and fourth, SD is within an almost equal range, here, the majority of the respondents are lying in SA (strongly agree). The major reason being the difference of opinion on this factor. As some people, observe a high level of awareness. However, others take in a different way i.e. generating a high level of waste.

TABLE IV. Causes of Waste Generation

No.	Cause of Waste Generation	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Standard Deviation	Relative Important index	Ranking
1	Increase Population	281	47	34	11	11	1.5	0.972	0.9	1
2	Lack of awareness among the public	103	126	55	56	44	2.51	1.33	0.6979	4
3	Poor Management	279	50	33	14	8	1.49	0.945	0.901	2
4	Restrict Open Dumping Yards	227	12	66	35	44	2.11	1.467	0.7786	3

Therefore, these products should not be discarded at open dumpsites, appropriate disposal methods such as incineration, etc. should be implemented. Fig.3 and Fig.4 demonstrate the open dumping of solid waste on the road.

The second significant factor, as found by this study, is public awareness. The most common problem of the day now is a lack of knowledge about waste. It has been found that the public is not interested in reducing the production of waste but is increasing it by any means. The people do not accept it as their responsibility to reduce waste and dump it on roads that would make the world unpleasant and unsafe, and then start criticizing the government. The main role the public itself plays in this regard is to make them aware of the need of the hour through various social activities.

Fig 1: Open dumping by the community of solid waste on the road.

Poor or bad management of higher authorities has also been stated to be a major cause and ranked second. Management is the key to success for any organization in any situation but mismanagement, whether residential or commercial, can fail. Likewise, inadequate management of restaurants, commercial areas, and construction work contributes to a rise in the solid waste that can be avoided by creating an accurate management plan for each organization.

Fig 3: Dumping sites on the road

Furthermore, insufficient awareness of how to store materials may cause pre-use material harm, i.e. proper use of containers and materials to minimize waste (e.g., creatively storing and using leftovers). It will help householders reduce household waste; especially it is necessary for Sanghar city because most of the city is a residential area.

Table V represents a Pearson Correlation Matrix showing the association of dependent and independent variables. The Asterisk (*) in the table shows that the 2- tailed correlation is significant at the 0.01 level, having a confidence level of 99%. In each cell represents its strength or relationship, whose detail is given in the legend of Table 4. The value of ±1 shows a perfect correlation, the value of ±0.9 to ±0.7 shows strong correlation, the value of ±0.6 to 0.4 shows moderate correlation, the value of ±0.3 to ±0.1 shows weak correlation, and unfilled cells shows no correlation. Most of the correlation values are negative; it shows how the relationship negatively affected the other variables.

The number of males was 57.3%, and the number of females was 42.7%. Overall, 34.6% of respondents were aged 18 to 35, whereas 13.5% were above 60 years. Regarding household size, 33.9% responded that they were 2 to 4 family members live in the house, 25.5% were 5 to 7 family members, 22.7% were 8 to 10 family members, while 18.0% replied that they were more than 10 family members in their house. Concerning education and income, about 32.3% had a post-graduation level of education, and 33.9% belonged to the category receiving a monthly income of RS 10,000 to 20,000 PKR, while 20.1% received RS 2000 to 10,000. About the profession, 33.6% were unemployed, and 2.6% was pensioner (see Table VI).

TABLE V. Correlations among variables

		1	2	3	4	5	6	7	8	9
1	Gender									
2	Age	.42**								
3	Education	-.64**	-.50**							
4	Employment	-.31**	-.42**	.27**						
5	Household size	.36**	.71**	-.47**	-.30**					
6	Income	-.59**	-.72**	.61**	.37**	-.63**				
7	Empty Dustbin	.52**	.81**	-.55**	-.39**	.60**	-.66**			
8	Generate waste	.51**	.75**	-.62**	-.34**	.69**	-.76**	.72**		
9	Disposal waste	-.44**	-.49**	.64**	.06	-.56**	.55**	-.51**	-.53**	
10	Waste collection	-.66**	-.51**	.78**	.20**	-.48**	.48**	-.50**	-.50**	.64**
	**. The correlation at level 0.01 (2-tailed) is significant. Note. N= 384, * $\rho < .05$, ** $\rho < .01$									

The description of the model and parameters for environmental factors are shown in Table VII. A value of 0.755 for 'R' shows that the environmental factors on municipal solid waste have a significant impact on people's quality of life. A value of 0.571 for 'R^2' shows that the predictors lie close to the line of regression, i.e., a strong prediction point. A value of 0.565 for 'Adjusted R^2' shows that 56.5% of residents realize that environmental factors have a significant impact on municipal solid waste generation. Whereas waste disposal was found to harm the environment and health quality with values of -0.155, respectively.

TABLE VI. Demographic characteristic.

Characteristics	Frequency	%
Gender		
Male	220	57.3
Female	164	42.7
Age		
18 to 35 Years	133	34.6
36 to 45 Years	120	31.3
46 to 60 Years	79	20.6
Above 60 Years	52	13.5
Occupation		
Employed	90	23.4
Unemployed	129	33.6
Self- Employed	56	14.6
Pensioner	10	2.6
Student	99	25.8
Education		
Uneducated	5	1.3
Primary	77	20.1
Matriculation	58	15.1
Intermediate	12	3.1
Graduation	108	28.1
Post-Graduation	124	32.3
Household-Size		
2 to 4	98	25.5
5 to 7	130	33.9
8 to 10	87	22.7
Above 10	69	18.0
Monthly Income		
RS 2000- RS 5000	77	20.1

RS 5001- RS 10,000	77	20.1
RS 10,000- RS 20,000	130	33.9
> 20,000	100	26.0

TABLE VII. Coefficient of regression for environment variables

R= 0.755, R^2= 0.571, Adjusted R^2= 0.565, Std. Error of Estimate= 0.683					
S. no	**Factor**	**Beta**	**Std Error**	**t-test**	**Sig.**
	Constant	1.587	0.198	8.036	0.000
X1	Empty Dustbin	0.418	0.029	14.662	0.000
X2	Disposal waste	-0.155	0.044	-3.490	0.001
X3	Waste collection	-0.076	0.042	-1.809	0.071
X4	Need for health education to create awareness regarding solid waste management	0.078	0.030	2.563	0.011
X5	Environmental impacts of solid waste around dumpsites	-0.00	0.027	-0.210	0.834

The overview model and parameter estimates for economic factors are shown in Table VIII. A value of 0.696 for 'R' shows that the economic factors influence all household sizes in a residential area significantly. A value of 0.484 for 'R^2' indicates the predictors lying on the regression line. A value of 0.479 for 'Adjusted R^2' indicates that 47.9 percent of residents think economic factors greatly influence the quality of life and the physical environment. Whereas income was found to be the negative impact factor with a coefficient value of -0.477. This can be explained as Sanghar City has gained a significant increase in residential density as a result of which the drainage and street conditions are getting worse and no one wants to pay for the cleaning of such waste that comes from the high-income family. They create more garbage as compare to the low-income group. Thus, people complain about the disposal system and the condition of the street.

TABLE VIII. Coefficient of regression for Economic variables

R= 0.696, R^2= 0.484, Adjusted R^2= 0.479, Std. Error of Estimate= 0.755					
S. no	**Factor**	**Beta**	**Std Error**	**t-test**	**Sig.**
	Constant	3.095	0.247	12.529	0.000
X6	Education	0.135	0.046	2.968	0.003
X7	Income	-0.477	0.047	-10.143	0.000
X8	How much will you pay?	0.212	0.040	5.298	0.000

Table IX displays the overview model and estimation of parameters for the rules and regulation factors. A value of 0.494 for 'R' shows that the rules and regulation factors are contributing less among effective and efficient solid waste management of restricted open_dumping_yards in Sanghar city. A value of 0.164 for 'R^2' indicates that the predictors are not outliers, but are far from the line of regression, i.e., not a good prediction point. A value of 0.156 for 'Adjusted R^2' shows that only 15.6% of improper waste management is caused due to the rules and regulation factors of solid waste. According to the estimation of the differences between actual and estimated variables by 'standard error of the estimate,' value of 1.058 indicates that the difference between actual and estimated variables is very high.

Whereas, increase penalties who violate the rules and regulations were found to be the element having a negative effect with a coefficient value of -0.100. And what the model predicts will not be as specific as the actual situation.

TABLE IX. Coefficient of regression for Rules and Regulation variables

R= 0.405, R^2= 0.164, Adjusted R^2= 0.156, Std. Error of Estimate= 1.058					
S. no	**Factor**	**Beta**	**Std Error**	**t**	**Sig**
	Constant	1.635	0.432	3.785	0.000
X9	Restrict_open_dumping_yards	0.304	0.037	8.131	0.000
X10	The management of SW is going to be solved.	0.304	0.171	1.773	0.077
X11	Increase penalties who violate the rules and regulations	-0.100	.041	-2.44	.015

In Sanghar city, there is a Lack of Planning and Management. There is no appropriate arrangement for the administration of SW. Therefore, the municipality accumulation of family waste is very irregular and constrained to powerful ranges. Therefore, strong waste stays scattered all through the rest of the areas. To distinguish the issues in MSW in Sanghar. To highlight the strong waste issues of the neighborhood individuals.

The current system of MSW administration at Sanghar city might be included with no appropriate principles and directions. Besides, this review can propose new and Improve instruments for the administration of civil strong waste administration. This exploration likewise can direct to receive a successful component of municipal solid waste administration to maintain a strategic distance from natural and well-being clashes in Sanghar city. Management and workers are not taking that waste the 86% of people are not satisfied with the actual situation of SW. The regression model used in this analysis to determine the environmental and public health effects is based on the variables selected. The main variables used in this work are descriptive statistics given in Table 6. The values depicted are based on the regular data points for all forms of MSW considered obtained from each source.

Taking into view the significance of a policy in solving urban issues, recommending policy implications was based on the results of the analysis. Economic, environmental, and rules and regulations factors were found as the cause of improper management. This results in achieving the study's objectives. To identify SWM issues through the inhabitant's perspective.

Fig.5 shows that the proposed plan for the municipal solid waste of Sanghar city and dumping site is also mentioned in this map. Green points show the skip containers and each container should be placed in every corner of the street, and the black arrow shows the dumping site, which is 50 acres of land, and this land is used for landfilling purposes.

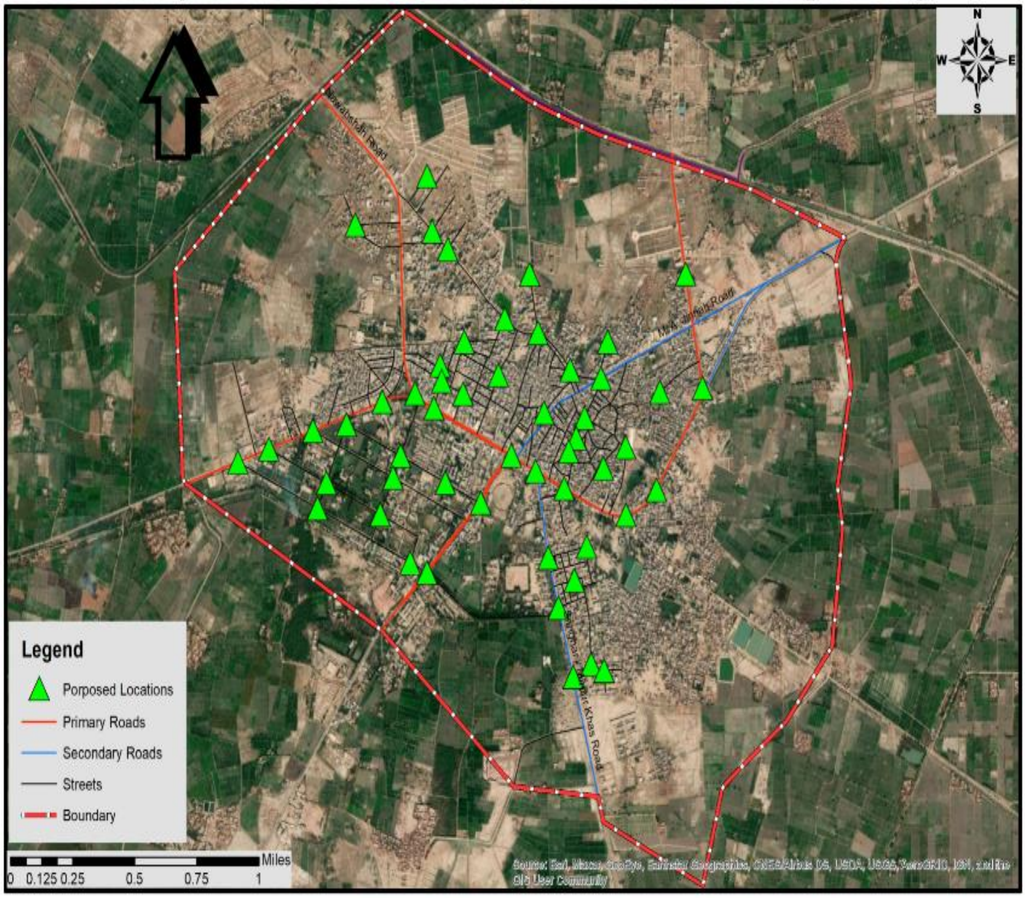


Fig 5: Image showing the proposed plan for skip containers and dumping site in Sanghar city.

5. Conclusion

My thesis aims to suggest a proposal for proper MSW dumping sites in Sanghar to enhance the city's quality of life and physical climate. The following goals are set to achieve this goal, and they are focused on data collected from primary and secondary sources.

Municipal solid waste management is essential for a city's growth and physical condition because improper solid waste disposal degrades human health and causes diseases in the environment. This problem, however, can be resolved through proper waste management and collection, which is a critical need in every region, including Sanghar, which has been chosen as the research area. The research examines several causes of SW generation in Sanghar through visits and a simple questionnaire survey. The level of significance of the causes was observed by collecting data between the different respondents to the survey questionnaire. The study concludes that, with population increases, a lack of public awareness, poor management are a few of the critical causes of waste generation in the city. Discussions with experts have revealed that controlling the population, including controlling urban growth in the city, can bring down major issues. Increasing public awareness, effective management of the various organizations can lead to a decrease in waste generation in the city. The study through the results suggests that there is an utter need for an efficient SW management system in Sanghar. In this regard, the Local Government needs to put their efforts. The overall system from collection to its safer disposal requires special attention. Moreover, public awareness can help the system more. The Local Government should not only utilize the public funds on the maintenance of the entire system properly but also requires focusing on arranging a few programs like public meetings, seminars, workshops, etc. on monthly basis. These programs would be aware of the public, which is an important stakeholder in the reduction and proper management of SW. The results show that the residents of Sanghar city have not facilitated to dispose of their waste properly daily. 60% of the Cumulative Percent of ashes was produced because of the open burning of MSW. 28% of SW is disposed into the open space and 31%on the dump side where openly burn the solid waste. 33%of solid waste disposes of residents by self and due to lack of management and workers are not taking that waste the 86% of people are not satisfied with the current situation of solid waste. The regression model used in this study to identify effects on the environment and public health is based on the variables selected. The main descriptive statistics variables used in this work are shown in Table 6. The values shown are based on the daily data points collected from each source for all MSW types considered. The study can be expanded further to equate related work with other major cities in Pakistan and abroad. In this way, thorough literature in different countries can be checked and potential approaches can be further refined.

AUTHOR CONTRIBUTION*:* All authors contribute equally in this research work.

DATA AVAILABILTY STATEMENT: The data of the manuscript will be available upon the request.

CONFLICT OF INTEREST: None

FUNDING: There is no any source of funding for this research.

References

[1] R. e. a. Rajkumar, "Effect of naked neck gene on immune competence, serum biochemical and carcass traits in chickens under a tropical climate," Asian-Australasian Journal of Animal Sciences, vol. 23, no. 7, pp. 867-872, 2010.

[2] G. e. a. Es'haghi, "Carbon nanotube reinforced hollow fiber solid/liquid phase microextraction: a novel extraction technique for the measurement of caffeic acid in Echinacea purpurea herbal extracts combined with high-performance liquid chromatography," Journal of Chromatography A, vol. 1217, no. 17, pp. 2768-2775, 2010.

[3] T. e. a. Tchobanoglous, Integrated solid waste management: Engineering principles and management lssues, McGraw-Hill, 1993.

[4] R. O. Bonow and L. A. e. a. Smaha, "World Heart Day 2002: the international burden of cardiovascular disease: responding to the emerging global epidemic," Circulation, vol. 106, no. 13, pp. 1602-1605, 2002.

[5] E. D. Anomanyo, "Integration of municipal solid waste management in Accra (Ghana): Bioreactor treatment technology as an integral part of the management process," Lund University, Sweden, 2004.

[6] H. I. Abdel-Shafy and M. S. Mansour, "Solid waste issue: Sources, composition, disposal, recycling, and valorization," Egyptian journal of petroleum, vol. 27, no. 4, pp. 1275-1290, 2018.

[7] R. E. Marshall and K. Farahbakhsh, "Systems approaches to integrated solid waste management in developing countries," Waste management, vol. 33, no. 4, pp. 988-1003, 2013.

[8] M. A. Akhund and H. U. e. a. Imad, "Risk attributes, influencing the time and cost overrun in joint venture construction projects of Pakistan," Engineering, Technology & Applied Science Research, vol. 8, no. 4, pp. 3260-3264, 2018.

[9] T. WaterAid, "Sanitation and water: Why we need a global framework for action," London: WaterAid, 2008.

[10] P. a. S. P. Grover, "An Analytical Study oif Effect of Family Income and Size on Per Capita Household Solid Waste Generation in Developing Countries," American Research Institute for Policy Development, vol. 3, no. 1, pp. 127-143, 2014.

[11] A. Dino and U. a. S. K. Mustafa, "MUNICIPAL SOLID WASTE MANAGEMENT-ANALYSIS OF WASTE GENERATION: A CASE STUDY OF RAWALPINDI CITY, PAKISTAN," A J o u r n a l o f A d v a n c e s i n M a n a g e m e n t , I T a n d S o c i a l S c i e n c e s, 2011.

[12] Y. a. T. Y. Dhokhikah, "Solid waste management in Asian developing countries: challenges and opportunities," Journal of Applied Environmental and Biological Sciences, vol. 2, no. 7, pp. 329-335, 2012.

[13] K. Dobraja and A. a. R. M. Barisa, "Cost-benefit analysis of integrated approach of waste and energy management," Energy Procedia, vol. 95, pp. 104-111, 2016.

[14] N. a. J. N. S. Ejaz, "Solid waste management issues in small towns of developing world: a case study of Taxila City," International Journal of Environmental Science and Development, vol. 3, no. 2, p. 167, 2012.

[15] A. R. Khoso and A. H. e. a. Memon, "Solid waste management issues in Hyderabad city," Mehran University Research Journal of Engineering and Technology, vol. 37, no. 3, pp. 653-662, 2018.

[16] L. A. Guerrero and G. a. H. W. Maas, "Solid waste management challenges for cities in developing countries," Waste management, vol. 33, no. 1, pp. 220-232, 2013.

[17] J. Chen and M. A. e. a. Hamon, "Solution properties of single-walled carbon nanotubes," Science, vol. 282, no. 5386, pp. 95-98, 1998.

[18] N. J. R. G. C. B. L. F. a. T. S. Turner, "From invisibility to transparency: identifying the implications," Ecology and society , vol. 13, no. 2, 2008.

[19] P. Pasunon, "Sample Size Determination from Krejcie and Morgan (1970) approach in quantitative research," The Journal of Faculty of Applied Arts, vol. 2, pp. 112-20, 2014.

[20] G. Israel, "Determining sample size: University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS; 1992," Fact Sheet PEOD-6, 2013.

[21] R. Gill, S. Kalwar, and I. A. Memon, "Yeh's Satisfaction Index Modelling of Tenants in Rental Appartments (A Case Study of Latifabad Hyderabad)," Sukkur IBA Journal of Computing and Mathematical Sciences, no. 2, pp. 1-10%V 4, 2021-01-05 2021, doi: 10.30537/sjcms.v4i2.654.

[22] S. Kalwar, I. A. Memon, and S. Qureshi, "Significance of National Spatial Planning for Economic Development of Secondary Cities in India: Critical Analysis of JNNURM Programme," Sukkur IBA Journal of Computing and Mathematical Sciences, no. 2, pp. 49-60%V 4, 2021-01-05 2021, doi: 10.30537/sjcms.v4i2.658.

[23] I. A. Memon, "Factors influencing travel behaviour and mode choice among Universiti Teknologi Malaysia employees," Universiti Teknologi Malaysia, Faculty of Built Environment, 2010.

[24] I. A. MEMON, "Mode Choice Modelling to Shift Car Travelers Towards Park and Ride Service in the CBD of Putrajaya and Karachi," PhD, Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, 2018. [Online]. Available: http://utpedia.utp.edu.my/id/eprint/18372

[25] I. A. Memon, S. Kalwar, N. Sahito, S. Qureshi, and N. Memon, "Average Index Modelling of Campus Safety and Walkability: The Case Study of University of Sindh," Sukkur IBA Journal of Computing and Mathematical Sciences, no. 1, pp. 37-44%V 4, 2020-07-29 2020, doi: 10.30537/sjcms.v4i1.582.

[26] I. A. Memon, N. Madzlan, M. A. H. Talpur, M. R. Hakro, and I. A. Chandio, "A review on the factors influencing the Park-and-Ride traffic management method," in Applied Mechanics and Materials, Kuala Lumpur, June 2014, vol. 567: Trans Tech Publications, Ltd., pp. 663-668, doi: https://doi.org/10.4028/www.scientific.net/AMM.567.663.

[27] I. A. Memon, M. Napiah, M. A. Hussain, and M. R. Hakro, "Influence of factors to shift private transport users to Park-and-Ride service in Putrajaya," in In Engineering

Challenges for Sustainable Future: Proceedings of the 3rd International Conference on Civil, Offshore and Environmental Engineering (ICCOEE 2016, Malaysia, 15-17 Aug 2016), N. A. W. A. Zawawi Ed., 1st Edition ed. London: CRC Press, 2016, p. 566.

[28] I. A. Memon, M. Napiah, M. A. H. Talpur, and M. R. Hakro, "Mode choice modelling method to shift car travelers towards Park and Ride service," ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 6, pp. 3677-3683, 2016 2016.

[29] K. Shaikh, A. Memon, I. A. Memon, Z. A. Laghari, and A. M. Memon, "Awareness regarding Coronavirus pandemic among the population of Sindh, Pakistan: A cross-sectional study," Sukkur IBA Journal of Computing and Mathematical Sciences, no. 1, pp. 28-36%V 4, 2020-07-29 2020, doi: 10.30537/sjcms.v4i1.573.

[30] M. A. Almomani, S. Basri, and A. R. Gilal, "Empirical study of software process improvement in Malaysian small and medium enterprises: The human aspects," Journal of Software: Evolution and Process, vol. 30, no. 10, p. e1953, 2018.

[31] A. R. Gilal, J. Jaafar, S. Basri, M. Omar, and A. Abro, "Impact of software team composition methodology on the personality preferences of Malaysian students," in 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 2016: IEEE, pp. 454-458.

[23] A. R. Gilal, J. Jaafar, M. Omar, S. Basri, and A. Waqas, "A rule-based model for software development team composition: Team leader role with personality types and gender classification," Information and Software Technology, vol. 74, pp. 105-113, 2016.

[33] A. R. Gilal, M. Omar, J. Jaafar, K. I. Sharif, A. Mahesar, and S. Basri, "Software development team composition: personality types of programmer and complex networks," in 6th International Conference on Computing and Informatics (ICOCI-2017), 2017, pp. 153-159.

[34] M. Z. Tunio, H. Luo, C. Wang, F. Zhao, A. R. Gilal, and W. Shao, "Task Assignment Model for Crowdsourcing Software Development: TAM," Journal of Information Processing Systems, vol. 14, no. 3, 2018.

Vol. 5, No. 1 | January – June 2021

A QR Code Based Group Pairing Approach for Mobile Ad Hoc Networks

Yasir Arfat Malkani[1], Moez Ahmed Malik[1], Lachhman Das Dhomeja[2], Abdul Waheed Mahessar[2], Bisharat Rasool Memon[2]

Abstract:

Due to the rapid growth of small and smart hand-held devices, mobile ad hoc networks (MANets) are becoming very common nowadays. MANets may consist of a number of small hand-held devices having limited resources in terms of memory, battery and processing power. In order to provide services to the users, these devices are capable of communicating with each other through some radio technology, such as WiFi, Bluetooth or Infrared. Since radio channels are inherently vulnerable to various security threats, it requires that devices in MANets must establish a secure association amongst themselves before exchanging any sensitive information or data. The process of establishing a secure channel between two devices is referred to as device pairing or device association. Device pairing do not rely on traditional mechanisms for security due to the impulsive and ad hoc interactions among the devices. Due to this, researchers have proposed many approaches to deal with this issue; however, the issue of group pairing (i.e. secure association of more than two devices) is less addressed issue in the literature yet. There could be many scenarios (such as confidential office meetings, paring of group of home appliances in smart-homes, etc) of MANets, where secure group communications is desired. Consequently, this research focuses on this issue and proposes a QR (quick response) code based approach to establish a secure channel between a numbers of devices. The proposed system is implemented and tested on modern hand-held devices and a usability study of the implemented system is also carried out

Keywords: *Group Pairing; Mobile Ad hoc Networks; Security; Device Association; QR Code*

1. Introduction

Mobile Ad hoc Networks (MANETs) are emerged from distributed computing and mobile computing [1]. In recent years, MANETs have revolutionized the computing world through their enormous useful applications in varying sub-fields. The main goal of mobile ad hoc network is to provide services to its users anytime and anywhere and to achieve this goal, devices need to connect with each other spontaneously. Due to the wireless nature of Mobile Ad hoc Networks, these are vulnerable or open to various security attacks [2 – 4]. Also note that in Mobile Ad hoc Networks devices do not share security credentials a priori, so traditional security mechanisms cannot be applied to Mobile Ad hoc Networks. This raises the need of bootstrapping the security process before actual data transfer. This security bootstrapping process is referred as secure first connect or device paring in the literature. The issue of device pairing is not new in the field of Mobile Ad hoc Networks. Many researchers

[1]Institute of Mathematics & Computer Science (IMCS), University of Sindh, Jamshoro
[2]Department of Inforamtion Technology, University of Sindh, Jamsoro
Corresponding Author: Yasir Arfat Malkani (yasir.malkani@usindh.edu.pk)

have worked on it and proposed many solutions to it during last two decades. However, most of them have focused on only two device scenarios and there has been less focus on group device paring. Group device pairing refers to establish a secure channel between more than two devices.

As stated, achieving the goal of secure first connect is non-trivial task in Mobile Ad hoc Networks due to its open nature. Not only this, but the conventional methods to handle Man-in-the-Middle (MiTM) and eavesdropping attacks are also inapplicable due to their computational cost and fixed infrastructure requirements. Due to this, researches proposed various modern solutions to solve the problem of device pairing based on the concept of out-of-band (OOB) channels.

Out-of-band (OOB) channel refers to an additional or secondary location constrained channel that provides some additional security properties to establish a secure channel at a short distance. Some examples of OOB channel includes near-field-communication (NFC), infrared (IrDA), audio and visual channels, etc. So far, researchers have given many solutions to device pairing using these OOB channels, but these solutions are mainly for two-device scenarios. In contrast to these solutions, authors in [2] proposed an approach to device pairing called PoP (Proof-of-Proximity) framework that have combined various device pairing protocols into a generic system for providing usable and secure scheme for a larger set of device pairing scenarios including two-device and multiple devices scenarios.

Though, PoP framework is providing a generic solution to device pairing and is easy to use, but still it has limited support for group pairing (i.e: associating more than two devices securely over a short range wireless channel). In this research work, it is advocated that there could be many scenarios (such as confidential office meetings, paring of group of home appliances in smart-homes, etc) of MANets, where secure group communications are desired. As a result, in contrast to previous approaches [5 – 29] to device pairing, this research focuses on the issue of group pairing and proposes a scheme to build a secure channel between group of devices.

1.1 Research Contributions

The research contributions of this paper are listed below:

- The first and main contribution of this research work is the proposal of a QR (quick response) code based approach to group device pairing.
- The second contribution of this work is the implementation of the proposed group device pairing scheme.
- The third contribution of this research work is the conduct of a usability study of the proposed solution. The usability study is carried out to verify whether the proposed solution is user-friendly/usable or not.

2. Background

In this section, the basic terminologies and concepts related to the field of device pairing are presented followed by a comprehensive and detailed literature survey of the device pairing schemes.

2.1 Mobile Ad Hoc Network

A mobile ad hoc network (MANET) is a infrastructure-less network of mobile devices connected wirelessly that have capability of self-configuration and self-organization. In other words, a mobile ad hoc network is consisting of numerous small hand-held devices having limited resources in terms of memory, battery and processing power [30], [31]. In order to provide services to the users, these devices need to communicate with each other through some radio technology, such as WiFi, Bluetooth or Infrared.

2.2 Device Pairing

Device pairing refers to setup a connection between two unassociated devices in a secure manner prior to exchanging any information or data using a short-range wireless technology, such as infrared, Bluetooth, WiFi, etc [24].

2.3 Device Pairing Protocol

Protocol refers to a set of rules governing the exchange of data or information over a communication channel and device pairing protocols refers to the approach or set of rules that are used to initiate a secure channel between two or more devices [2].

2.4 Out-of-Band Channel

In the literature of device pairing, out-of-band (OOB) or side channel refers to a secondary communication channel that is used to exchange some minimal security material (such as short password or random number) to initiate the pairing process. Some examples of OOB channels include audible voice, LED lights, bar codes, NFC technology, etc [2].

2.5 Literature Review

In this section a compressive literature review on state-of-the-art device pairing schemes is presented. This literature review is divided into two sub-sections: the first sub-section describes the pairing schemes that are mainly proposed for establishing secure channel between two devices, while the second sub-section discusses the device pairing schemes, which could be used in group scenarios and allow to connect more than two devices securely.

2.5.1 Device Pairing Schemes for Two-Devices Scenarios

Due to the significance of device pairing in both mobile ad hoc and ubiquitous computing environments, there is an immense research work done on this topic and during the last two decades many device paring schemes and protocols have been proposed. In the literature of device pairing, the work done by Stajano and Aderson [20], [22] is considered as the first that attracted other researchers towards this domain. They [22] presented a policy-based mother-duckling (i.e. resurrecting duckling) security model, which uses plain-text to transfer security material (i.e. encryption key) over physical medium (i.e. wire) to establish the secure channel between the devices. The drawback of this approach is that it is difficult to carry cables all the time and exchange of encryption key in plain-text is vulnerable to dictionary attack [32]. Another limitation of this approach is that it requires that both devices should have similar type of physical interface/port.

Later on Blafanz et al. [19] extended the work of Stajano and Aderson and proposed the first wireless channel based solution to device pairing. They [19] used Infrared as an out-of-band channel to exchange the cryptographic material between two devices to initiate the pairing process. The limitation of this approach is that it requires line-of-sight between the two devices while performing the pairing process. Some other approaches similar to Balfanz et al. approach are also proposed [24], which use the laser or ultrasound channels as out-of-band channels to exchange the minimal cryptographic material to initiate the pairing process.

The device association approaches or mechanisms that are discussed above use wired and/or proximity constrained channels, however, some researchers have proposed device association approaches, which use sensors. These approaches mainly are based on the concept of shaking devices together. In this category, the very first approach is Smart-its-Friend [21]. Later on this approach [21] is modified by Lester et. al. [18], which is called Are You With Me. Are You With ME is followed by another similar approach [10] that uses the same concept of shaking devices together. Shake-Well-Before-Use [10] require from the user to shake the two devices simultaneously in order to pair them. Note that these schemes use accelerometers data to bootstrap the pairing process. The main drawback of these approaches is that practically it is not applicable to shake the two devices simultaneously all the time due to their large size or being fixed in ceilings or at walls. Another similar approach [16] is Shake-Them-Up, which instead of accelerometer data exploits radio signals to bootstrap the secure device association process.

Later on AMIGO [11] - a radio based approach to device pairing - is proposed by Varshavsky et al.. They in fact extended the traditional Diffie-Hellman [32] key exchange approach for secure device association. Since the proposed approach exploits the WiFi access points data to pair the devices, it is not applicable to scenarios, where there is no WiFi (wireless fidelity) data is accessible to process.

Afterwards, Biometric technology is used as a location-limited side channel for device pairing. Biometric based OOB channels are combined with standard cryptographic primitives to accomplish the goal of secure first connect (i.e. device pairing). Biometric based approaches to device pairing are more attractive solutions, because they put little or no cognitive load/overhead on users [6], [8], [33]. In contrast to benefits of biometric based solutions, these also have some serious limitations. For example, these approaches need exhaustive calculations for pattern matching. Some researchers also proposed device pairing approaches based on NFC technology [12], [34]. NFC based approaches exploit magnetic field induction mechanism to bootstrap the secure pairing process. However, it is to be noted that NFC is an extremely short range technology among all other technologies that decreases its chances of applicability in device association scenarios where there devices are kept at some distance and also note that NFC is open to various attacks, such as data modification, corruption and eavesdropping [8].

Recently, some researchers have also proposed device pairing schemes, which are based on audio/video out-of-band channels. For example, See-is-Believing (SiB) [26] is one of them which uses camera and bar codes for device paring, while another approach called Loud-and-Clear (LaC) [14] uses microphone/speaker and display to establish the secure connection between two devices. SiB is inappropriate solution for the devices that do not have camera, while LaC is not a suitable solution for hearing-impaired users. HAPADEP is another audio based approach proposed by Sorient et al. [35]. The limitation of this approach [35] is that if users do not carefully listen to audio generated by devices, then the devices security may easily be compromised by false match. Apart from above approaches, some standard pairing technique, such as Bluetooth pairing [36], are also in operation, which uses PIN or password to connect the two devices securely, however, PIN code number is also vulnerable to exhaustive search attack and/or dictionary attack. By launching these attacks a four (04) digits PIN code could be broken down in less than 0.06 sec [37], [38]. Another limitation of Bluetooth pairing is that it requires human intervention to input the same PIN code / short password on both the devices to bootstrap the secure device association process.

While comparing all above discussed approaches with the approach proposed by Malkani [2], it is very clear that this is more standard and generic solution to device pairing. Malkani [2] called his developed approach as the proof-of-proximity (PoP) framework. The proposed approach [2] uses the common device capabilities, a discovery mechanism and several integrated device association schemes to provide a comprehensive support for a larger set of device association scenarios ranging from two-device settings to multiple device settings. The discovery system exploits or combines the best features of service location protocol (SLP) [39] and Universal Plug and Play (UPnP) [40] protocol. The readers, who are interested to know more about this approach can refer [2] for further reading.

In short, many device pairing approaches have been introduced by several researchers so far and each of them have their own advantages and disadvantages. Also note that above all approaches (excluding the [2], which has partial support for group pairing scenarios) are specially designed for pairing of two devices scenario. In next section, those pairing approaches are discussed, which are proposed pairing of more than two devices and thus are suitable for group pairing scenarios.

2.5.2 Device Pairing Schemes for Group Pairing Scenarios

As stated earlier in introduction section, there is very less work done on group pairing and to the best of my knowledge I have found only two direct approaches to group pairing. The first one is proposed by Ming Li et al. [41] and second more recent approach to group pairing is proposed by Zhiping Jiang et al. [42]. Ming Li et al. [41] focused on the secure sensors' association problem in body area network (BAN). All the sensors are securely paired before the body area network is actually deployed. This group pairing approach is based on traditional symmetric-key cryptographic primitives. The body area network is developed to meet a larger range of applications (e.g. ubiquitous health monitoring system (UHM) [43] and emergency medical response system (EMS) [44]).

Zhiping Jiang et al. [42] proposed an approach - called NFV - to allow a group of modern hand-held devices equipped with a motion sensor to exchange data and information securely. In their proposed scheme a group of people or users put their hand-held devices on a table and wait for the group connection to be established. The proposed approach exploits vibration technology to establish the secure channel.

2.6 Justification of the Proposed Group Pairing Approach

As stated mobile ad hoc networks (MANets) are becoming common day by day. A MANetis consisting of a number of small hand-held devices having capability of interacting with each other wirelessly. Since wireless technology is inherently vulnerable to several security threats (such as eavesdropping, main-in-the-middle (MiTM) attack), it requires some mechanisms to provide secure communication between the devices. As stated in previous section, during the last two decades, many research efforts [20-45] have been made that address the issue of security in general and security of Mobile Ad hoc Networks (MANETs) in particular. Each of the solutions have their own trade-offs in terms of device heterogeneity, usability and applicability. From the literature survey, it is concluded that researchers have proposed and developed many scheme to secure the communication between any two devices, however the issue of secure group association is less addressed yet. We advocate that there could be many MANets scenarios where there is a need of secure group communication. Therefore, in this research we propose to develop some mechanism that facilitates to bootstrap the secure pairing process between a group of devices. The proposed approach is also compared with one of the recent approach [42] to group pairing. The comparative analysis is presented in section 5. The results show that the proposed approach to group pairing is more effective as compared to the prior [42] group pairing scheme.

3. The Proposed System

The main goal of this research work is to develop a device pairing scheme that allows the secure association of group of devices (i.e. two or more devices) in a mobile ad hoc network. Therefore, in subsequent sections, the high-level and low-level design of the proposed approach is presented followed by a message sequence diagram, which summarizes the overall approach.

3.1 High-level Design of the Proposed System

Figure-1 illustrates the high-level design of the proposed approach. In the proposed approach there are two types of roles for devices. The first role is of master device or group controller, while the other role is of an ordinary group member. The difference between these two roles is that apart from being used as an ordinary member, a group controller is also responsible for initiating a group pairing process by distributing the initial key through QR code. Note that in the proposed approach QR code serves the purpose of out-of-band (OOB) channel. The brief description of the high-level design is given below:

It is assumed that the mobile phones intended to become the part of secure group communication must have cameras to read the QR code and are also having support for wireless interface to connect and exchange data or information with each other in a group setting.

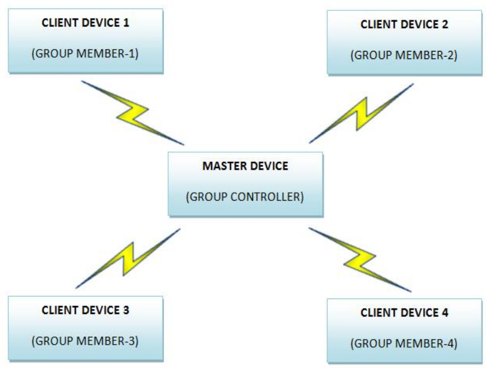


Figure-1: High-level architecture of the proposed system

In the proposed approach the master device first encodes some cryptographic material (i.e. a random short secret key) into a QR code and display it on its screen. Then, the user of every client device intended to become member of the group approaches the master device and reads the QR code (through the camera of his/her device) displayed on the master device to get the initial short secret key. Once the short secret key is obtained by the member device, it uses that short secret key to exchange the long term symmetric key in encrypted mode with the group controller through the normal in-band (i.e. WiFi) channel for further communication. Once the symmetric key is exchanged between the member device and group controller, it guarantees the secure communication between both of these devices. Consequently, later on group controller shares a common group key with group members in a secure way that ultimately results in secure group communication. The low-level and more technical discussion on the proposed approach is presented in subsequent sections.

3.2 Low-level Design of the Proposed System

In this section, the details of the proposed algorithm for group device pairing is presented preceded by the notations that could help in understanding the proposed algorithm.

Algorithmic Notations:

MD: Master Device (Group Controller)

GMD: Group Member Device

GCIK: Group Communication Initiation Key

MSG: Message

OTP: One Time Password

SGCK: Shared Group Communication Key

Step-1: (a) MD (Master Device) --> Generates GCIK (Group Communication Initiation Key)
(b) MD --> Generates SGCK (Shared Group Communication Key)
Step-2: MD --> Set MSG1:=Encode(GCIK)$_{QRCode}$
Step-3: MD --> Display(MSG1)
Step-4: GMD (Group Member Device) --> Read(MSG1)$_{MD}$
Step-5: GMD - -> Set GCIK:= Decode(MSG1)$_{MD}$
Step-6: GMD--> Generate OTP (One Time Password)
Step-7: GMD --> Set MSG2:=Encrypt(OTP)$_{GCIK}$
Step-8: GMD --> Send(MSG2) to MD
Step -9: MD --> Set OTP$_{GMD}$:= Decrypt(MSG2)$_{GCIK}$
Step-10: MD --> Set MSG3:=Encrypt(SGCK) with OTP$_{GMD}$
Step-11: MD --> Send(MSG3) to GMD
Step-12: GMD --> Set SGCK := Decrypt(MSG3)
[Repeat steps 4 to 12 for every device intended to become part of the group communication]
Step-13: Return

Figure-2: The proposed algorithm for secure group device association

In figure-2, the proposed algorithm for secure group device pairing is presented. The description of each of the step of the algorithm is presented below:

In very first step, the master device (i.e. group controller or manager) generates a group communication initiation key (GCIK) that is a short secret key used to initiate the pairing process and a long term symmetric key, called shared group communication key (SGCK) that is actually used for group communication in a secure mode. Once the GCIK and SGCK are generated, the master device encodes GCIK only into a quick response (QR) code, while keeps the SGCK with itself. QR code (i.e an array of black and white squares) is a special type of bar-code that could easily be read by mobile phones or other digital devices. QR codes are usually used for encoding short or limited information that could be read by any camera enabled device. After the encoding of GCIK within the QR code, it is displayed on the screen of master device.

Once the GCIK is displayed on the screen of master device, any other device/mobile phone intended to become the part of group communication need to access this QR code first. In this regard, the user of the member device need to approach the master device and read the QR code through the camera of member device. Then, the member device decodes the QR code to obtain the GCIK. After obtaining the GCIK, the member device generates its own one-time-password (OTP) key and encrypts it with GCIK and send it back to the master device through ordinary in-band wireless channel for further communication between them in a confidential/encrypted mode. This process (i.e. step - 4 to step - 12) is repeated for each device that intended to become the part of group communication.

Note that once SGCK is shared, the group pairing is achieved and hence the group member devices could securely communicate with each other using standard symmetric cryptographic primitives.

3.3 Message Sequence Diagram

In order to understand the proposed group pairing approach at high-level, a message sequence diagram is presented in figure-3. This message sequence diagram itself is self-explanatory and shows all the same steps that are explained in previous sections.

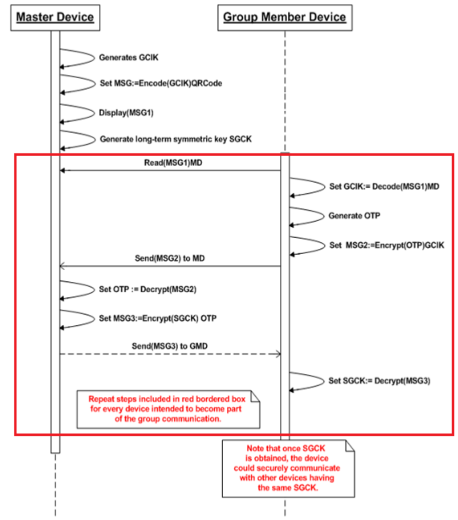


Figure-3: Message sequence diagram of the proposed approach

4. System Implementation and Testing

In order to validate or verify the viability of a theoretical solution, it is always desired to implement the system in real word and test it. Consequently, the proposed system is also implemented and tested in real-world scenarios.

The proposed approach/algorithm is implemented and tested using a Laptop that works as a WiFi hotspot and mobile phones running the Android operating system and coding is done using Android studio. Following mobile phones are used in implementation and testing phase:

- Samsung J5 (android 5.1 / 1.2 GHz quad core / 1.5 GB RAM)
- Samsung Mega (android 4.2 / 1.7 GHz dual core / 1.5 GB RAM)
- Qmobile Qtab x50 (android 4.2.2/1.2 Quad core / 1 GB RAM)

The reason to choose Android based mobile phones is that Android is an open source operating system and most of the modern mobile phones support it. Android studio IDE, as shown in figure-4, is installed on a Corei7 machine with 8GB RAM with 3.6 GHz processing power and having Windows 7 operating system running on it.

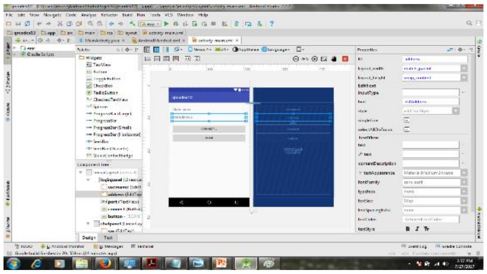


Figure-4: Android studio IDE

4.2 System Testing

In this section the working mechanism of the proposed system is demonstrated through some real-world test scenarios, which are prepared for testing the implementation of the proposed system.

The figures 5 to 8 show the snapshots taken during the testing of the proposed system after successful implementation. Figure-5 shows the Laptop that is used as WiFi hotspot and three mobile phones in which middle one is the master device (i.e. group controller) and other two are clients (i.e. group member devices).

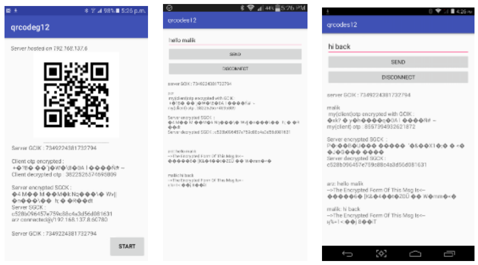
Figure-5: Device involved in system testing

Figure-6: Client 1 reading QR code to obtain GCIK

Figure-6 illustrates the situation when client1 (i.e. group member device) scans the QR code from the master device in order to obtain the GCIK. Client1 once receives the GCIK, it generates its OTP and sends it to the master device encrypted with GCIK through WiFi channel.

Figure-7 shows the situation when client2 (i.e. group member device) scans the QR code from the master device in order to obtain the GCIK. Client2 once receives the GCIK, it generates its OTP and sends it to the master device encrypted with GCIK through WiFi channel.

Figure-7: Client 2 reading QR code to obtain GCIK



(a) (b) (c)

Figure-8: Sample of logs generated on each device involved in pairing process
(a) Master Device (Group Controller)
(b) Client-1 (Group Member Device)
(c) Client-2 (Group Member Device)

Figure-8 shows the three snapshots taken after successful pairing of group of devices. Figure-8 (a) shows the log generated at master device during the pairing process, while the Figure-8 (b) and Figure-8 (c) show log/text generated on client 1 and client 2 devices respectively. It could be seen in figures 8 (b) and 8 (c) that both of the devices are exchanging information/data between themselves through a chat application, which is developed to test the secure communication between group members after successful pairing. This testing procedure is repeated several times to make sure that the system working well and is reliable. After system testing, a usability study of the developed system is performed and the details of that study are discussed in next section (i.e. results and discussion) in detail.

5. Results and Discussion

In order to meet the aims and objectives of this research work, we proposed a group device association scheme for mobile ad hoc networks (MANETs). The proposed scheme is implemented, tested and its usability evaluation is also performed to confirm that the proposed system is user-friendly and practically a usable solution for group device association scenarios.

The usability evaluation test equipment are the same as used during system implementation and testing. The only difference is that during the usability testing, we allowed the test participants to use their own smart phones so that more realistic working of the proposed system could be analyzed. In the usability evaluation seventy two (72) participants are recruited to use the developed system and provide their feedback about it and also rate the developed system on 7-point likert-type-scale [45], [46].

Table-1: Usability evaluation participants' demographic data

	Nos.	%age
Gender:		
Male	57	79%
Female	15	21%
Age:		
18 – 23	52	72%
24 – 28	15	21%
29 or above	5	7%
Last Qualification:		
Intermediate	16	22%
BS/MSc/MCS or Equivalent Degree	46	64%
MS/M.Phil	8	11%
PhD or above	2	3%
Occupation:		
Teaching	6	8%
Student	58	81%
Other	8	11%

As stated earlier, the usability evaluation of the developed system is conducted by recruiting seventy two (72) participants. The usability evaluation participants are either students, teachers or employees of University of Sindh, Mehran University of Engineering & Technology and Liaquat University of Medical and Health Sciences, Jamshoro. The demographic data of the participants is provided in table-1, which is self-explanatory and requires no more description.

5.1 Usability Evaluation Results

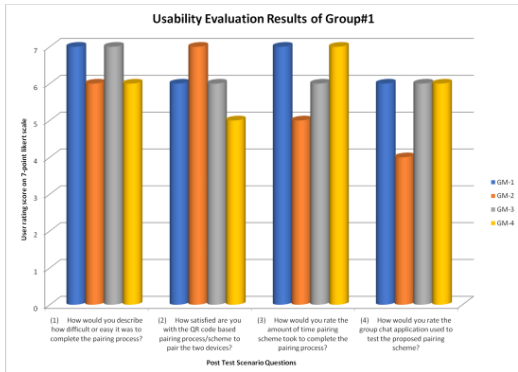
This section presents the results that are obtained by analyzing the data gathered during the usability evaluation. In order to collect the usability data, two (02) questionnaires (i.e. pre-test questionnaire and post-test questionnaire) are used. In pre-test questionnaire usability evaluation participants are asked to provide their demographic information, while post-test questionnaire is used to record the participants rating scores for four (04) usability evaluation questions with regard to the developed system. The user rating is carried out through seven-point likert scale in which each participant is asked to provide their ratings by selecting the score in the range 1 to 7. One is considered as the least score, while seven indicates the highest or the most satisfactory score. Microsoft Excel package is used to store and process the collected data for analysis purposes.

Figure-9: Usability evaluation participants while using the developed system in small groups (i.e. 4 members in each group)

Figure-10: Usability evaluation participants while using the developed system in larger groups (i.e. 10 members in group)

Figures 9 and 10 show some random snapshots taken during the usability study. In usability study, total fourteen (14) groups are formed consisting of overall seventy two (72) members. First twelve (12) groups are consisting of four (04) members in each, while the 13th group is consisting of ten (10) members and the 14th group is consisting of fourteen (14) participants. The reason to increase the numbers in group 13 and 14 is to test whether the proposed solution is workable in scenarios where number of users increase and decrease time-to-time.

The charts or graphs in figure 11 to figure 16 show the results of usability evaluation for first ten (10) groups having four (04) participants in each group. Every usability participant is provided four different questions through the post-test scenario questionnaire and asked to provide the rating for each of them.



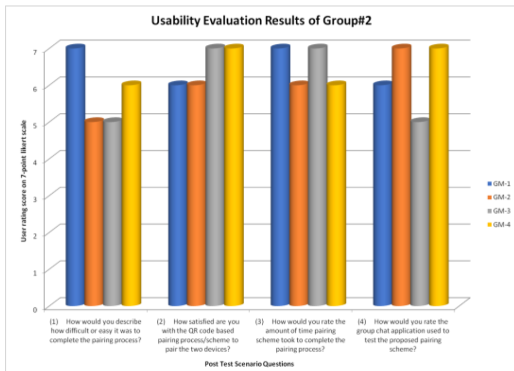


Figure-11: Usability evaluation scores as given by participants of groups #1 and #2

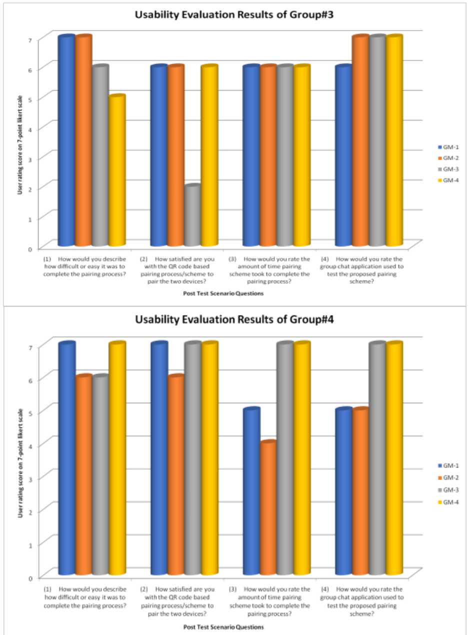


Figure-12: Usability evaluation scores as given by participants of groups #3 and #4

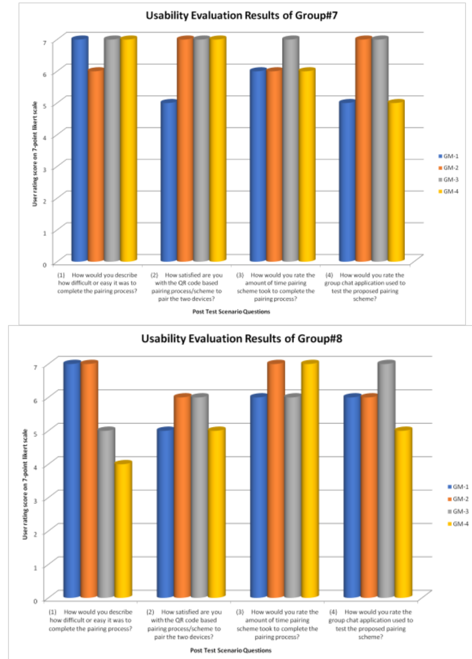


Figure-14: Usability evaluation scores as given by participants of groups #7 and #8

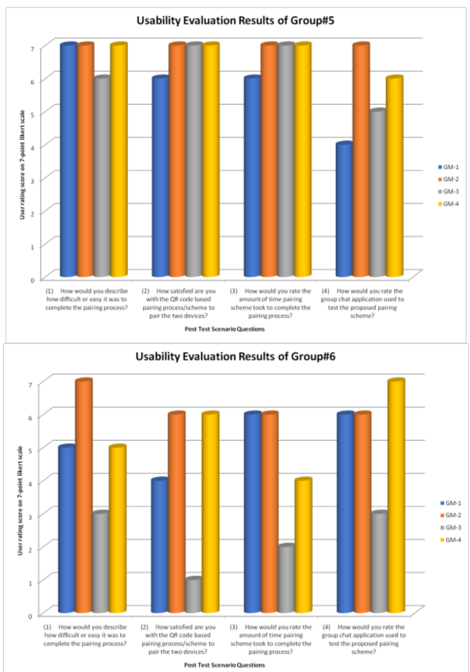


Figure-13: Usability evaluation scores as given by participants of groups #5 and #6

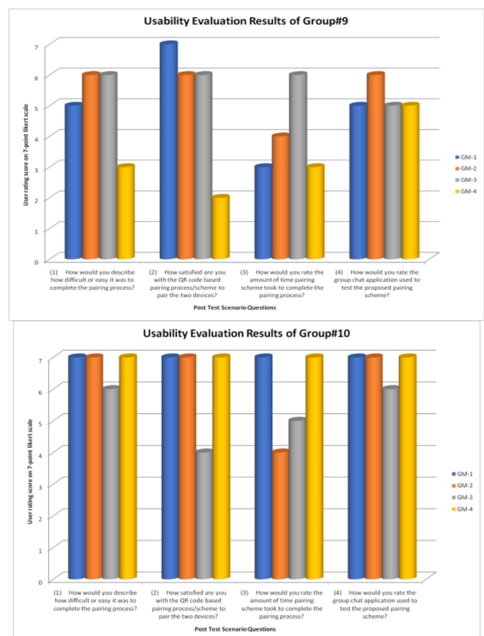


Figure-15: Usability evaluation scores as given by participants of groups #9 and #10

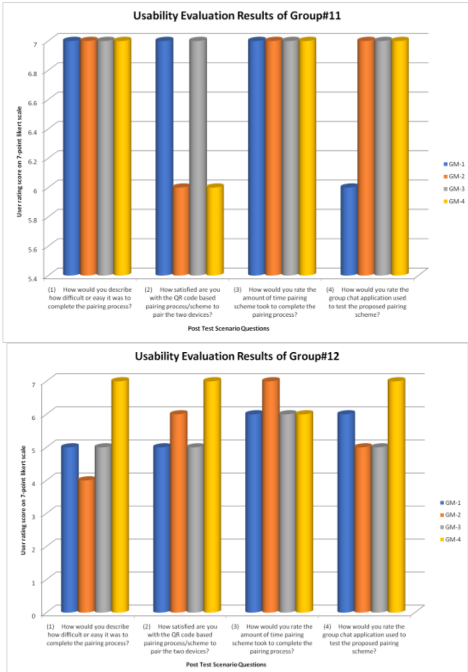


Figure-16: Usability evaluation scores as given by participants of groups #11 and #12

The charts or graphs in figures 17 and 18 show the results of usability evaluation for last two groups.

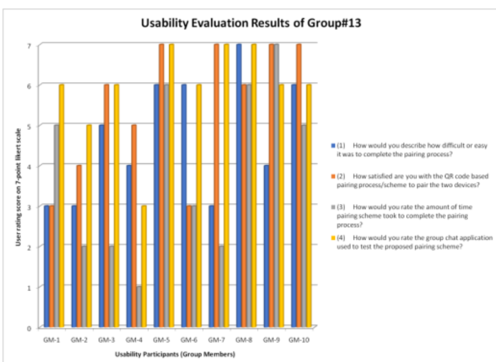


Figure17: Usability evaluation scores as given by participants of groups #13

Figure-17 refers to the results of group 13, where ten (10) participants performed the task of group pairing and provided their usability score as per their experience. Similarly, figure-18 shows the results obtained from group 14, which is consisting of fourteen (14) group members. Again in this setting, every usability participant is provided four different questions through the post-test scenario questionnaire and asked to provide the rating for each of them.

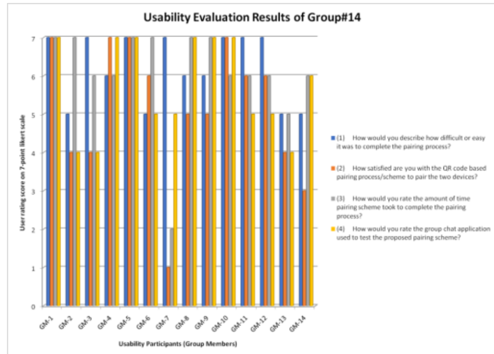


Figure-18: Usability evaluation scores as given by participants of groups #14

The chart shown in figure-19 provides the overall summary of the usability evaluation results. This chart shows the mean and standard deviation of all scores as provided by all of usability participants and as discussed/shown in figures 11 to 18.

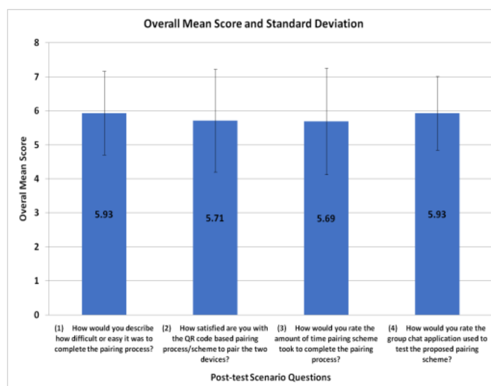


Figure-19: Overall mean and standard deviation of all the scores as given by usability evaluation participants

Nielsen. et al. [45] has indicated that while performing usability evaluation experiments through a seven-point likert scale, if one gets the mean score that is equals to 5.6 or above it, the produce or system under evaluation could be considered as usable and acceptable for use. Note that the last chart (figure 5.11) shows that for each of the usability question/parameter, the mean score is above 5.6. Consequently, the developed system is considered as one of the good and usable solution for the scenarios where there is need of associating more thatn two devices in mobile ad hoc networks (MANETs).

5.2 Comparative Analysis

This section presents the comparative analysis of the proposed developed system with prior work on group device association. As stated in literature review section, there is very less work done on group devices assocation and hence we found only one work [42] that is closely relevant to the proposed system.

Table-2: Comparative analysis of the proposed approach with prior work on group device paring

Approach to group device association	Maximum supported distance between devices while initialing pairing	Out-of-Band (OOB) channel used
NFV: Near Field Vibration Based Group Device Pairing (Zhiping Jiang et al., 2015) [12]	Upto 40 cm (15.748 inches)	Vibrator and motion/vibrator detector sensor
The Proposed Approach	Upto 50.8 cm (20 inches)	QR (quick response) code and camera

Consequently, the proposed system is compared with it [42] and the comprasion results are shown in table 5.4. From table 5.4, it could be seen that the prposed systme is more better in terms of distance requied for initiating the pairing process. Also note that the work proposed by Zhiping Jiang et al. best performs only when they use plastic table / surface [42], however in our proposed approach there is no any restriction of surface type is imposed.

6. Conclusion

Mobile Ad hoc Networks (MANETs) are becoming common day by day. Devices in a MANET interact with each other through a wireless channel, which requires secure association between the communicating partners prior to exchanging any information or data. The concept of establishing secure association between two or more devices in close proximity is known as device paring. One of the less addressed aspect of device pairing is the association of more than two devices (i.e. also known as group pairing). In this paper, a solution to the issue of group pairing in MANets is presented. The outcome of the proposed research can be used in various scenarios of MANets where secure and spontaneous interaction among a group of devices is desired, such as smart conference halls and meeting rooms, etc. The most important and major contribution of this work is the proposal and development of a QR code based system that allows the pairing of a group of devices. The other two contributions of this research work are (a): a comprehensive and detailed literature survey of device pairing that could help the newbie researchers to understand the device paring domain in a very quick manner, and (b) the conduct of a usability evaluation of the proposed solution to verify whether the proposed solution is user-friendly or not. The usability evaluation results were also positive and the results indicated that the proposed system is working very well in real world scenarios and is one of the usable and acceptable solution from users point of view.

AUTHOR CONTRIBUTION*:* In this paper, Yasir Arfat Malkani and Lachman Das Dhomeja conceived the idea of the paper. Further, Yasir Arfat Malkani and Moez Ahmed Malik jointly performed the system Implementation and conducted usability study with some support from Lachhman Das Dhomeja. Abdul Waheed Mahesar and Bisharat Rasool Memon helped in literature survey and to refine/improve the system and also supported in writing the paper.

DATA AVAILABILTY STATEMENT: There is no any third party data is used in this research. However, the raw (usability study) data collected as part of this research is available (if required).

CONFLICT OF INTEREST: None

FUNDING: No any funding received and used for this research.

REFRENCES

[1] Y. A. Malkani and L. Das Dhomeja, "Secure device association for ad hoc and ubiquitous computing environments", International Conference on Emerging Technologies, Islamabad, Pakistan, 2009, pp. 437-442, doi: 10.1109/ICET.2009.5353132..

[2] Y. A. Malkani, “A proof-of-proximity framework for device pairing in ubiquitous computing environments,” A Univ. Sussex DPhil thesis, 2011.

[3] M. Qabulio, Y. A. Malkani, and A. Keerio, “On Node Replication Attack in Wireless Sensor Networks,” vol. 34, no. 4, pp. 413–424, 2015.

[4] M. Qabulio, Y. A. Malkani, and A. Keerio, “A framework for securing mobile wireless sensor networks against physical attacks”, 2016 International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 2016, pp. 1-6, doi: 10.1109/ICET.2016.7813265.

[5] N. Saxena, Md. Borhan Uddin, and J. Voris. 2008. Universal device pairing using an auxiliary device. In Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS'08). ACM, New York, NY, 56--67. DOI:http://dx.doi.org/ 10.1145/1408664.1408672

[6] I. Buhan, J. Doumen, P. Hartel, and R. Veldhuis, “Feeling is Believing : a location limited channel based on grip pattern biometrics and cryptanalysis,” no. 2.

[7] N. Saxena and J. Voris, “Pairing devices with good quality output interfaces,” Proc. - Int. Conf. Distrib. Comput. Syst., pp. 382–387, 2008.

[8] I. Buhan, J. Doumen, P. Hartel, and R. Veldhuis, “Secure Ad-hoc Pairing with Biometrics: SAfE,” Proc. 1st Int. Work. Secur. Spontaneous Interact. (IWSSI ’07), pp. 450–456, 2007.

[9] C. Soriente, G. Tsudik, and E. Uzun, “BEDA: Button-enabled device association,” Int. Work. Secur. Spontaneous Interact., 2007.

[10] R. Mayrhofer and H. Gellersen, “Shake Well Before Use: Authentication Based on Accelerometer Data,” Pervasive Comput., pp. 144–161, 2007.

[11] A. Varshavsky, A. Scannell, A. LaMarca, and E. de Lara, “Amigo: Proximity-Based Authentication of Mobile Devices,” pp. 253–270, 2007.

[12] E. Haselsteiner and K. Breitfuß, “Security in Near Field Communication (NFC) Strengths and Weaknesses,” Semiconductors, vol. 11, no. 71, p. 71, 2006.

[13] BTSIG, “Simple Pairing Whitepaper,” Bluetooth Spec. Interes. Gr., vol. 1, no. 1, p. 23, 2006.

[14] M. T. Goodrich, M. Sirivianos, J. Solis, G. Tsudik, and E. Uzun, “Loud and Clear : Human-Verifiable Authentication Based on Audio 1 Introduction 2 Related Work,” Elements, pp. 1–16, 2006.

[15] N. Saxena, J. Ekberg, K. Kostiainen, and N. Asokan, “Secure Device Pairing based on a Visual Channel (Short Paper) ∗,” pp. 2–7, 2006.

[16] C. Castelluccia, “Shake Them Up!,” ACM/Usenix Mobisys, pp. 51–64, 2005.

[17] A. Pathan, Security of self-organizing networks: MANET, WSN, WMN, VANET. 2016.

[18] J. Lester, B. Hannaford, and G. Borriello, “‘Are You with Me?’ – Using Accelerometers to Determine If Two Devices Are Carried by the Same Person,” pp. 33–50, 2004.

[19] D. Balfanz, D. K. Smetters, P. Stewart, and H. C. Wong, “Talking to Strangers: Authentication in Ad Hoc Wireless Network,” Proc. 9th Netw. Distrib. Syst. Secur. Symp., pp. 46–56, 2002.

[20] F. Stajano, “The resurrecting duckling—what next?,” Secur. Protoc., pp. 204–214, 2001.

[21] L. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl5, and H.-W. Gellersen, “Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts,” Ubicomp 2001 Ubiquitous Comput. SE - 10, vol. 2201, pp. 116–122, 2001.

[22] F. Stajano and R. J. Anderson, “The Resurrecting Duckling: Security Issues for Ad-hoc Wireless Networks,” Proc. 7th Int. Work. Secur. Protoc., pp. 172–194, 2000.

[23] M. K. Chong, R. Mayrhofer, and H. Gellersen,

"A survey of user interaction for spontaneous device association," ACM Comput. Surv., vol. 47, no. 1, p. 8, 2014.

[24] S. Mirzadeh, H. Cruickshank, and R. Tafazolli, "Secure device pairing: A survey," IEEE Commun. Surv. Tutorials, vol. 16, no. 1, pp. 17–40, 2014.

[25] Y. A. Malkani, "A Generic Framework for Device Pairing in Ubiquitous Computing Environments," Int. J. Netw. Secur. Its Appl., vol. 4, no. 2, pp. 1–20, 2012.

[26] J. M. McCune, A. Perrig, and M. K. Reiter, "Seeing-Is-Believing: using camera phones for human-verifiable authentication," Int. J. Secur. Networks, vol. 4, no. 1/2, p. 43, 2009.

[27] C. Soriente, G. Tsudik, and E. Uzun, "Secure pairing of interface constrained devices," Int. J. Secur. Networks, vol. 4, no. June, p. 9, 2009.

[28] R. Prasad and N. Saxena, "Efficient device pairing using 'human-comparable' synchronized audiovisual patterns," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5037 LNCS, pp. 328–345, 2008.

[29] N. Saxena and M. B. Uddin, "Automated device pairing for asymmetric pairing scenarios," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5308 LNCS, pp. 311–327, 2008.

[30] J. Loo, J. Mauri, and J. Ortiz, Mobile ad hoc networks: current status and future trends. 2016.

[31] S. Al-Sultan, M. Al-Doori, … A. A.-B.-J. of network and, and undefined 2014, "A comprehensive survey on vehicular ad hoc network," Elsevier.

[32] E. Hellman, "New Directions in Cryptography," 1976.

[33] R. M and A. Jain, "Biometrics : The Future of Identification It is too early to predict where , how , and in which form reliable biometric," Technology, pp. 46–49, 2000.

[34] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis, "Touch {&} connect and touch {&} select: interacting with a computer by touching it with a mobile phone," MobileHCI '09 Proc. 11th Int. Conf. Human-Computer Interact. with Mob. Devices Serv., p. 1, 2009.

[35] C. Soriente, G. Tsudik, and E. Uzun, "HAPADEP: Human-Assisted Pure Audio Device Pairing," in Information Security, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 385–400.

[36] D. Sangster, C. Dreher, … J. N.-U. P., and undefined 2016, "Systems and methods for pairing bluetooth devices," Google Patents.

[37] C. Soriente, G. Tsudik, and E. Uzun, "Information Security," vol. 5222, no. September 2008, 2008.

[38] J. P.-N. S. Publication and undefined 2017, "Guide to bluetooth security," nvlpubs.nist.gov.

[39] E. G.-I. I. Computing and undefined 1999, "Service location protocol: Automatic discovery of IP network services," ieeexplore.ieee.org.

[40] 278 DY Cheng - US Patent App. 09/742 and undefined 2002, "UPnP enabling device for heterogeneous networks of slave devices," Google Patents.

[41] L. Ming, Y. Shucheng, L. Wenjing, and R. Kui, "Group Device Pairing based Secure Sensor Association and Key Management for Body Area Networks BT - INFOCOM, 2010 Proceedings IEEE," pp. 1–9, 2010.

[42] Z. Jiang, J. Han, W. Xi, and J. Zhao, "NFV: Near Field Vibration Based Group Device Pairing," 2016, pp. 129–140.

[43] E. Jovanov, A. Milenkovic, C. Otto, and P. C. De Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," vol. 10, pp. 1–10, 2005.

[44] K. Lorincz et al., "Sensor Networks for Emergency Response: Challenges and Opportunities," Pervasive Comput., vol. 3, no. 4, pp. 16–23, 2004.

[45] Xiaopeng Li, Qiang Zeng, Lannan Luo, and Tongbo Luo. 2020. "T2Pair: Secure and Usable Pairing for Heterogeneous IoT Devices". In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20). Session 1E: Cyber Physical Systems November 9–13, 2020, New York, NY, USA, pg: 309–323.

[46] J. Nielsen, J. L.-C. of the ACM, and undefined 1994, "Measuring usability: preference vs. performance," dl.acm.org.

[47] A. Seffah, M. Donyaee, R. B. Kline, and H. K.

Padda, “Usability measurement and metrics: A consolidated model,” Softw. Qual. J., vol. 14, no. 2, pp. 159–178, Jun. 2006.

Sukkur IBA Journal of Computing and Mathematical Sciences

SUKKUR IBA UNIVERSITY
Merit - Quality - Excellence

SUKKUR IBA UNIVERSITY
Airport Road, Sukkur -65200
Sindh, Pakistan
Tel: +92-71-5644233
Fax: +9271-5804419
Email: sjcms@iba-suk.edu.pk
URL: sjcms.iba-suk.edu.pk

Sukkur IBA Journal of Computing and Mathematical Sciences
SJCMS

Sukkur IBA Journal of Computing and Mathematical Sciences

Recognized by HEC Pakistan in “Y” Category

E-ISSN: 2522-3003
P-ISSN: 2520-0755

Volume: 5 | No: 2 | Jul - Dec | 2021