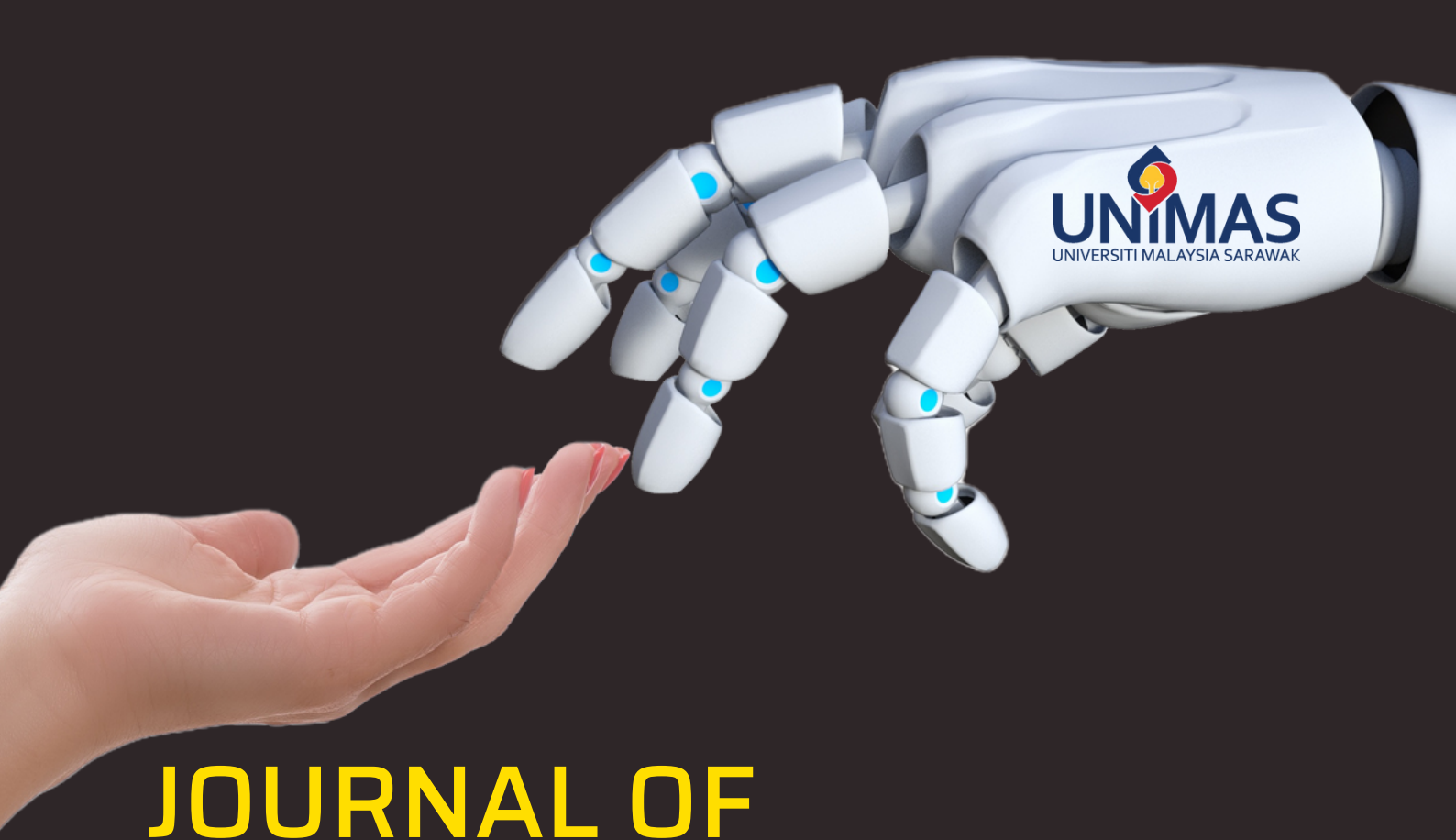


JOURNAL OF COMPUTING AND SOCIAL INFORMATICS

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITI MALAYSIA SARAWAK

eISSN 2821-3777

Editorial Committee

Advisory Board

Reviewers

Journal of Computing and Social Informatics

The Journal of Computing and Social Informatics (JCSI) is an international peer-reviewed publication that focuses on the emerging areas of Computer Science and the overarching impact of technologies on all aspects of our life at societal level. This journal serves as a platform to promote the exchange of ideas with researchers around the world.

Articles can be submitted via *www.jcsi.unimas.my*

Assoc Prof Dr Chiew Kang Leng

Chief Editor
Journal of Computing and Social Informatics
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia

Contents

Evaluating the Generalizability of Support Vector Machine for Breast Cancer Detection

[1]Oluwaseyi Ezekiel Olorunshola, [2*]Okeh Dominic Ebuka and [3]Adeniran Kolade Ademuwagun

Department of Computer Science, Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria

email: [1]seyisola25@yahoo.com, [2*]okehoroko2019@email.com, [3]kademuwagun@gmail.com

**Corresponding author*

Received: 22 July 2024 | Accepted: 19 November 2024 | Early access: 26 November 2024

Abstract - *Breast cancer is caused by abnormal cell growth in the breast. Early detection has been observed to be crucial for successful treatment. Accurate detection methods are essential. Machine learning models, particularly Support Vector Machines (SVMs), have shown promise. However, concerns exist regarding their generalizability across real-world scenarios with varying software environments and data processing techniques. This research investigates this gap by comparing SVM performance with other classifiers such as Naïve Bayes, Random Forest, Multilayer Perceptron and Decision Tree. These classifiers were tested on the Wisconsin Breast Cancer dataset using both the Waikato Environment for Knowledge Analysis (WEKA) and Jupyter Notebook. The study recorded performance metrics such as accuracy, precision, recall, and f1_score. After the analysis, it was observed that in WEKA, Support Vector Machine under the 10-fold cross-validation and 70% split, had the highest accuracies of 0.981 and 0.977 respectively. Interestingly, Multilayer Perceptron also achieved an accuracy of 0.977 under the 70% split. In the Jupyter Notebook, Support Vector Machine also produced the highest accuracy value of 0.99 under the 70% split. However, Random Forest produced the highest accuracy of 0.97 which was closely followed by Support Vector Machine which had a value of 0.96 in the 10-fold cross-validation.*

Keywords: Malignant, Benign, Support Vector Machine, Classifiers, Evaluation Metrics, Breast Cancer.

1 Introduction

Breast cancer occurs when abnormal cells in the breast grow uncontrollably, forming tumors that can potentially spread throughout the body and become life-threatening (WHO, 2023). The year 2020 saw 2.3 million women diagnosed with breast cancer and 685,000 deaths worldwide. By the end of 2020, 7.8 million women who had been diagnosed with breast cancer in the previous 5 years were still alive, making it the most common cancer globally (WHO, 2023). From recent research, it was stated that timely detection of the disease can lead to a positive prognosis and a high chance of survival. In North America, patients with breast cancer have a 5-year relative survival rate of over 80% due to the early detection of the disease (Sun et al., 2017). The traditional method for detecting cancer relies on a gold-standard approach involving three tests: radiological imaging, clinical examination, and pathology testing. This conventional method relies on regression to determine the presence of cancer. The effective incorporation of Information and Communication Technologies (ICT) into medical practice has become a crucial factor in the modernization of the healthcare system, particularly in the realm of cancer treatment (Naji et al., 2021). The latest machine learning (ML) techniques and algorithms are developed based on model design. ML is a computational approach that can be used to find the best solutions to a problem without requiring explicit programming by a computer programmer or an experimenter (Akbuğday, 2019). The utilization of ML models, particularly Support Vector Machines (SVMs), has displayed notable potential in the realm of breast cancer detection through the analysis of mammograms and other imaging modalities. However, concerns exist regarding the generalizability of these models when applied in real-world settings. Current research on SVM models for breast cancer detection often evaluates them in single environments and with limited variations in training and testing methodologies. This raises questions about whether the reported accuracy and precision translate well to different software platforms and data splitting techniques. This research aims to investigate the generalizability of SVM models for breast cancer detection by comparing its performance with other classifiers such as Naïve Bayes (NB), Random Forest (RF), Multilayer Perceptron (MLP) Neural Network (NN), and

Decision Tree (DT) under various programming environments and data splitting techniques. The performance of each classifier will be compared across both environments and data-splitting techniques using the chosen evaluation metrics. Statistical tests will be conducted to assess the significance of any observed differences. This research hypothesizes that while SVM models might achieve high accuracy in specific environments, their performance may deteriorate when applied to different software platforms or with alternative data-splitting methods. This research is expected to reveal the generalizability of SVM models for breast cancer detection. By comparing SVM with other algorithms under various conditions, the study will provide valuable insights into the robustness and reliability of these models in practical applications. The analysis of this research was carried out using the Waikato Environment for Knowledge Analysis (WEKA), a tool that provides Classification, Clustering, Association Mining, Feature Selection, and Data Visualization (Shah & Jivani, 2013). Additionally, Python's Jupyter Notebook was used to evaluate the performance of these five classifiers using the four performance metrics: accuracy, precision, recall, and F1-score. This was done to validate the results obtained from WEKA. The remaining part of this paper is arranged as follows; Section 2 contains literature review while Section 3 contains the methodology. Results are discussed and analyzed in Section 4 while Section 5 concludes the paper.

2 Literature Review

Hoque et al. (2024) utilized the Extreme Gradient Boosting (XGBoost) ML technique to detect and analyze breast cancer. The breast cancer Wisconsin (diagnostic) dataset was used in the study and it comprised of 569 rows, where each row denoted a distinct digitized image of a breast mass and 33 columns. Out of 569 rows, no column had missing data besides the "Unnamed: 32" column which only had null values. It was stated in the study that in contrast to linear regression models, ML models like XGBoost and RF models were generally resistant to multicollinearity between features. Hence, for this problem, the researchers refrained from using a linear regression model. The result stated that XGBoost provided an accuracy of 94.74% and a recall of 95.24%.

Islam et al. (2024) evaluated and compared the classification accuracy, precision, recall, and F1-scores of five different ML methods using a primary dataset (500 patients from Dhaka Medical College Hospital). It was stated that ML and Explainable Artificial Intelligence (AI) were crucial in classification as they not only provide accurate predictions but also offered insights into how the model arrived at its decisions, aiding in the understanding and trustworthiness of the classification results. Five different supervised ML techniques, including DT, RF, logistic regression (LR), NB and XGBoost, were used to achieve optimal results on the dataset. The study applied SHAP analysis on the XGBoost model to interpret the model's predictions and understand the impact of each feature on the model's output. After the final evaluation, the XGBoost achieved the best model accuracy score, which was 97%.

Dinesh et al. (2024) carried out a study to compare the efficacy of the state-of-the-art SVM method for image prediction with that of K-Nearest Neighbors (KNN), LR, RF, and DT. The study made use of the UCI ML Laboratory which provided a total of 569 samples. The maximum acceptable error was set at 0.5, and the minimum power of analysis was set at 0.8. Predictions made using LR appeared to have a higher accuracy (95%) than those made using SVM, KNN, DT, or RF (92%, 90%, 85%, and 91%). This proposed system had a probability importance of 0.55.

Elsadig et al. (2023) selected eight classification algorithms that had been used to predict breast cancer to be under investigation. These classifiers include single and ensemble classifiers. A trusted dataset has been enhanced by applying five different feature selection methods to pick up only weighted features and neglect others. Accordingly, a dataset of only 17 features was developed, SVM is ranked at the top by obtaining an accuracy of 97.7% with classification errors of 0.029, False Negative (FN) and 0.019 False Positive (FP). Therefore, it was noteworthy that SVM was the best classifier and outperformed even the stack classier.

Using the Wisconsin Breast Cancer Diagnosis Dataset, Strelcenia and Prakoonwit (2023) presented an effective feature engineering method to extract and modify features from data and the effects it has on different classifiers. The feature was used to compare six popular ML models for classification. The models compared were LR, RF, DT, KNN, MLP, and XGBoost. The results showed that the DT model, when applied to the proposed feature engineering, was the best performing, achieving an average accuracy of 98.64%.

Chaurasiya and Rajak (2022) carried out an experiment to compare the accuracy measures of four prominent classification models considering their performance qualitatively on Wisconsin Diagnostic Breast Cancer (WDBC) dataset. RF, SVM, KNN and LR ML algorithms were analyzed on a classification technique that generally contains two different steps. In the first step the training dataset which contains labelled classes was used to build classification model by selecting a suitable classification algorithm. In the later step which is predictive phase, the accuracy of the built classification model was evaluated on the validation dataset. RF

classifier was experimentally observed to be the best algorithm with accuracy of 95% and precision of 90.9% as compared to the other three classifiers.

Guleria et al. (2020) research was based on the prediction and diagnosis of the classes of breast cancer (benign or malignant) by using supervised learning techniques in WEKA. The research made use of KNN (83.41% precision, 90.04% recall, 80.42% accuracy, and 0.86 F-Measure), NB (88.37% precision, 94.53% recall, 87.41% accuracy, and 0.91 F-Measure), LR (81.65% precision, 88.55% recall, 77.97% accuracy, and 0.84 F-Measure), and DT (85.71% precision, 92.53% recall,83.91% accuracy, and 0.88 F-Measure). It was observed that the prediction model built-up by NB provided the higher accuracy as well as higher F-measure among all the algorithm.

Ibeni et al. (2019) made use of three classifiers NB, BN, and Tree Augmented Naïve Bayes (TAN). The paper presented the fully Bayesian approach to assess the predictive distribution of all classes using three datasets: breast cancer, breast cancer Wisconsin, and breast tissue dataset. The prediction accuracies of Bayesian approaches were also compared with K-NN, DT (J48) and SVM. The result of the performance metrics evaluated on the algorithms were arranged according to accuracy, precision, recall, and F-measure for the breast cancer dataset Algorithm: KNN: 94.992%, 96.94%, 95.483%, and 96.207%. SVM: 96.852%, 97.161%, 98.017% and 98.591%. DT (J48): 94.992%, 95.633%, 96.688% and 96.157%. BN: 97.28%, 96.506%, 99.325% and 97.895%. NB: 95.994 %, 95.196%, 98.642%, and 96.888%. TAN: 96.280%, 95.851%, 98.430%, and 97.123%. The result showed that BN was the best performing algorithm.

Akbuğday (2019) investigated the accuracies of three different ML algorithms; k-NN, NB, and SVM using WEKA. The values of the report were as follows; K-NN had 96.85% accuracy, NB had 95.99% accuracy and C-SVM a sub-classifier of SVM had 96.85%. It was observed that K-NN and SVM algorithms were the most accurate ones with identical confusion metrics and accuracy values.

Keleş (2019) research was aimed at the prediction and detection of breast cancer early with non-invasive and painless methods that use data mining algorithms. In this study, an antenna was designed to operate in the 3-12 GHz UWB frequency range and a 3D breast structure consisting of skin layer, fat layer, and fibro glandular layer was designed. A separate model was also designed by adding a tumor layer to the breast structure. The dataset that was created had 6006 rows/values, 5405 of which were used as the training dataset, while 601 were used as the test dataset. The dataset was then converted to the arff format, which was the file type used by the WEKA tool. The 10-fold cross-validation technique was then used to obtain the most accurate results using the Knowledge Extraction based on Evolutionary Learning (KEEL) data mining software tool. The results indicated that Bagging, IBk, Random Committee, RF, and Simple CART algorithms were the most successful algorithms, with over 90% accuracy in detection.

3 Methodology

The research design, environment and dataset are described in this Section. And also, the algorithm and performance metrics are also examined. Two different environments were used to analyze the datasets in order to determine the best performing classifier out of the 5 classifiers against the 4 performance metrics for breast cancer prediction. These environments are WEKA and Python's Juypiter Notebook. The 10-fold cross-validation and 70% split was carried out in each of the 2 environments.

The following methods were used to compare WEKA's user-friendly platform, which is great for initial exploration and rapid prototyping, with Python's power and flexibility for building and deploying advanced ML models for breast cancer analysis. Each environment has its advantages; for instance, Python offers advanced techniques, scalability, integration, deployment, and sharing, whereas WEKA provides rapid prototyping, testing, and data processing tools. Akbuğday (2019) stated that due to WEKA's Java-based nature and comprehensive built-in algorithm library, employing the use of another platform with better-implemented algorithms environments such as Python or R may lead to more accurate classifiers with better programming practices and platform-specific advantages. Hence, this research was carried out using the 2 environments.

3.1 Research Design

This research utilizes a quantitative research methodology to conduct a comparative analysis of various ML algorithms, assessing their performance using specific metrics, to predict breast cancer mortality.

3.2 Environment Description

The analysis in this study was conducted using WEKA version 3.8.6 and Python Jupyter Notebook version 7.1.2. WEKA, developed by Holmes, Donkin, and Witten in 1994, is an open-source ML software that offers a comprehensive collection of tools for data preprocessing, classification, regression, clustering, association rules mining, and visualization. Jupyter Notebook is a project Spun off IPython in 2014 by Fernando Perez and Brian Granger. It is a non-profit, open-source project born out of the IPython in 2014 as it evolved to support interactive data science and scientific computing across all programming language. Jupyter Notebook was created based on Python Programming Language developed by Guido van Rossum, a Dutch programmer in the late 1980s.

3.3 Data Description and Preprocessing

The Breast Cancer Wisconsin (Diagnostic) dataset used in this study was sourced from the University of California Irvine (UCI) Repository. It comprises features extracted from digitized Fine Needle Aspirate (FNA) biopsies images. This dataset which consists of clinical and demographic features of breast cancer patients is a multivariate dataset which consists of 569 instances and 33 features and it has 0 mismatches and 0 missing values. For the Python environment, the dataset was loaded into the pandas DataFrame. This is crucial for making the data accessible for analysis and preprocessing. The data was examined to identify and remove any unintended unnamed columns that might exist due to formatting issues. With the dataset loaded and cleaned, the target variable, 'diagnosis,' was converted from categorical to numerical values. This was done to make the data compatible with ML algorithms. Specifically, the diagnosis labels 'M' (malignant) and 'B' (benign) were mapped to 1 and 0 respectively. The features were then scaled to ensure that they had a mean of 0 and a standard deviation of 1. This standardization is essential for ML models as it ensures that all features contribute equally to the model training process thereby improving convergence speed and overall performance. For the WEKA environment, WEKA automatically distinguishes between nominal, numeric and string attributes, and converts the selected column that comprises the target variables to the required format. WEKA automatically prompts users by applying scaling filters if it detects that an algorithm needs data in a specific range, ensuring compatibility without manual adjustment.

3.4 Performance Metrics

i. Accuracy: the ratio between the correctly classified samples and the total number of samples in the evaluation dataset. (Hicks et al., 2022).

$$ACCURACY = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

ii. The Recall: also known as the sensitivity or True Positive Rate (TPR), and is calculated as the ratio between correctly classified positive samples and all samples assigned to the positive class (Hicks et al., 2022.).

$$RECALL = \frac{TP}{TP + FN} \tag{2}$$

iii. The Precision: is calculated as the ratio between correctly classified samples and all samples assigned to that class. (Hicks et al., 2022).

$$PRECISION = \frac{TP}{TP + FP} \tag{3}$$

iv. F1-Score: The F1 score is the harmonic mean of precision and recall, meaning that it penalizes extreme values of either. (Hicks et al., 2022).

$$F1 = 2 * \frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

4 Result and Discussion

This Section discusses and analyzes the values of the performance metrics obtained on each classifier when 10-fold cross-validation and 70% split was applied on the dataset in both WEKA and Jupyter Notebook environments. Table 1 and Figures 1 to 4 show the results from WEKA while Table 2 and Figures 5 to 8 show the results from Python's Jupyter Notebook for accuracy, recall, F1-score and precision.

4.1 Discussion on WEKA Environment

The evaluation performed on the WEKA environment revealed a close competition between SVM and MLP in classifying breast cancer. Both models achieved impressive accuracy scores, with SVM reaching 0.981 under 10-fold cross-validation (as seen in Figure 1) and MLP achieving 0.977 under the 70% split (Figure 1). While SVM edged out MLP in terms of accuracy under cross-validation (Figure 1), MLP demonstrated a slight advantage in recall (0.975) under the 70% split (Figure 2). However, SVM maintained a consistently high F1-score (0.98 and 0.975) across both evaluation methods (Figure 3), indicating a good balance between precision and recall. Precision, as shown in Figure 4, also favored SVM with the highest values (0.983 and 0.978). This suggests SVM might be slightly better at correctly identifying true positives (cancerous cases). Based on these results, SVM appears to be the slightly better model for overall breast cancer prediction. It consistently achieved high performance across all metrics (accuracy, recall, F1-score, and precision) under both evaluation methods (Figures 1-4). However, MLP's strong performance, particularly in recall under the 70% split (Figure 2), suggests it could be a viable alternative depending on the specific needs of the application.

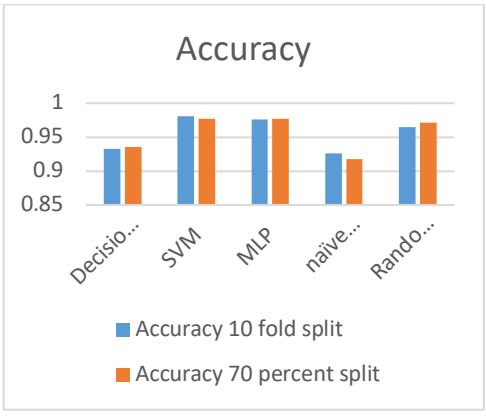


Figure 1: Comparison of accuracy in WEKA

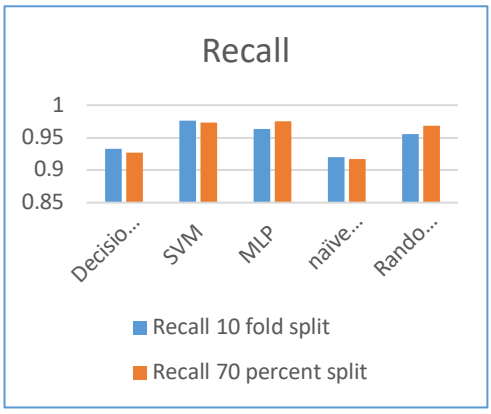


Figure 2: Comparison of recall in WEKA

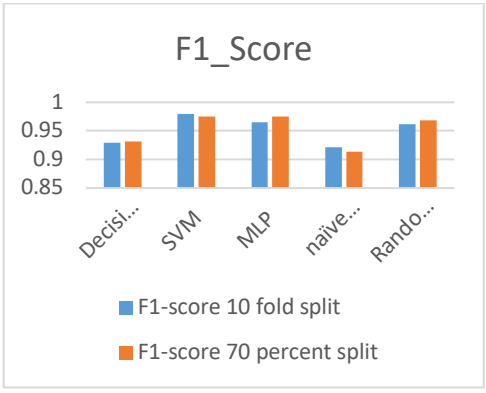


Figure 3: Comparison of F1_score in WEKA

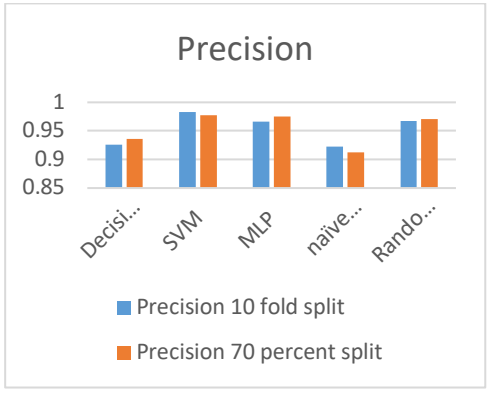


Figure 4: Comparison of precision in WEKA

4.2 Discussion on Python's Jupyter Notebook Environment

The python's Jupyter Notebook evaluation revealed an interesting dynamic between SVM and RF for breast cancer classification. While both models performed well, their strengths lie in different evaluation scenarios. On the 70% split (Figure 5), SVM excelled in accuracy, achieving a remarkable value of 0.99. However, under 10-fold cross-validation (Figure 5), RF emerged as the leader with an accuracy of 0.97. A similar pattern emerges in

recall (Figures 6 and 7). SVM dominated the 70% split with a recall of 0.99 (Figure 6), while RF led the 10-fold cross-validation with a score of 0.97 (Figure 7). This suggests that SVM might be slightly better at identifying true positives in a specific, pre-defined training/testing split, but RF might generalize better across unseen data. Precision analysis (Figure 8) follows the same trend. SVM reached a value of 0.99 in the 70% split, while RF achieved the highest score (0.97) under 10-fold cross-validation.

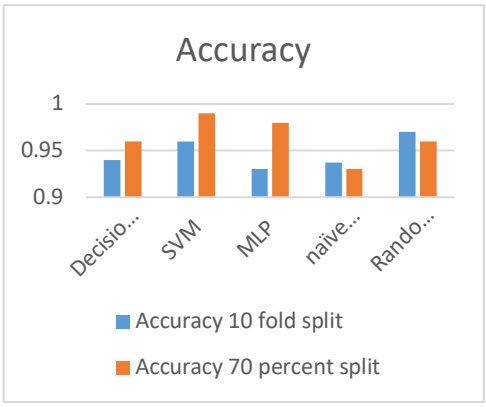


Figure 5: Comparison of accuracy in Python

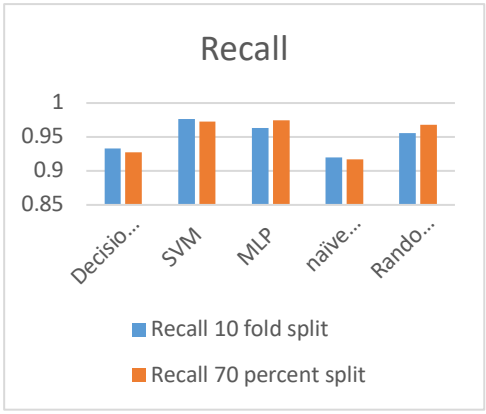


Figure 6: Comparison of recall in Python

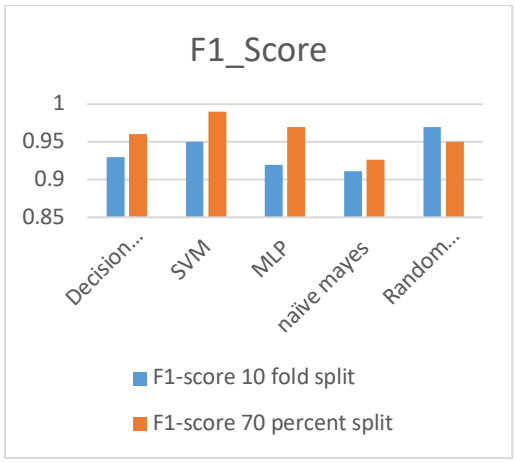


Figure 7: Comparison of F1_score in Python

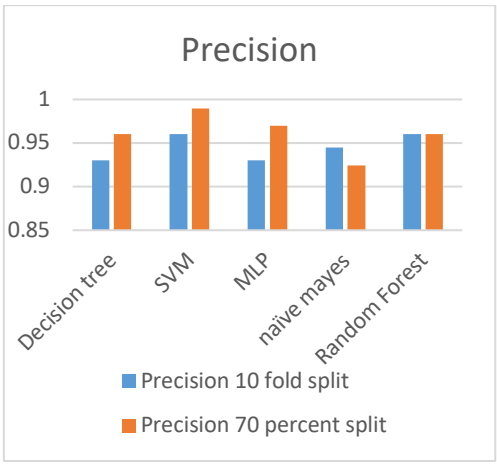


Figure 8: Comparison of precision in Python

4.3 Summary of Results

In summary, the analysis of the results shows that the performance of SVM can be influenced by the evaluation method or environment used to develop the model. It may perform well in accuracy on a specific training/testing split (e.g., 70% split) in one environment because of the 70% split allowing the model to learn more patterns and nuances which help SVM build a stronger decision boundary, but show lower performance under cross-validation in a different environment, because the 10-fold validation provides only a small training set in each fold limiting SVM ability to learn patterns effectively, which aims for better generalizability. SVM is a powerful algorithm for classification, especially when there is a clear margin of separation between classes. This makes it work well with high-dimensional spaces, like the Wisconsin dataset, which includes many features. SVM can however be sensitive to noise, particularly in datasets where classes overlap significantly, as the margin can become less meaningful. RF is an ensemble method that combines multiple DT to improve generalization and reduce overfitting. It performs well with imbalanced and complex datasets by averaging the predictions of individual trees. In cross-validation settings, RF often shows good generalization across folds. This is due to its ability to capture complex relationships in the data without relying on a single model structure. RF also handles outliers better and is less sensitive to overfitting compared to single DT or models that depend on a small number of support vectors. It is important to note that SVM and RF are the best-performing classifiers because the results obtained from both environments are consistent, unlike the MLP which performed well in WEKA but inconsistently in the Jupyter notebook, suggesting possible overfitting. Choosing between SVM and RF depends on the specific application and the importance of performance in a particular evaluation setting. If a well-defined

dataset is split and high accuracy on that specific data is the priority, SVM might be a good choice. However, if generalizability across unseen data is crucial, RF might be more suitable due to its stronger performance under cross-validation. Tables 1 and 2 show the summary of all the values gotten from the analysis of the dataset under the 10-fold cross-validation and 70% split in the WEKA and Jupyter Notebook environments.

Table 1: Result of the performance metrics of the algorithms in WEKA

	Accuracy		**Recall**		**F1-Score**		**Precision**	
	10-fold split	**70 % split**	**10-fold split**	**70 % split**	**10-fold split**	**70 % split**	**10-fold split**	**70 % split**
DT	0.933	0.935	0.933	0.927	0.929	0.931	0.926	0.936
SVM	**0.981**	**0.977**	**0.976**	0.973	**0.98**	**0.975**	**0.983**	**0.978**
MLP	0.976	**0.977**	0.963	**0.975**	0.965	**0.975**	0.966	0.975
NB	0.926	0.918	0.92	0.917	0.921	0.914	0.922	0.912
R F	0.965	0.971	0.956	0.968	0.962	0.969	0.967	0.971

Table 2: Result of the performance metrics of the algorithms in Python

	Accuracy		**Recall**		**F1-Score**		**Precision**	
	10-fold split	**70 % split**	**10-fold split**	**70 % split**	**10-fold split**	**70 % split**	**10-fold split**	**70 % split**
DT	0.94	0.96	0.93	0.96	0.93	0.96	0.93	0.96
SVM	0.96	**0.99**	0.95	**0.99**	0.95	**0.99**	**0.96**	**0.99**
MLP	0.93	0.98	0.92	0.97	0.92	0.97	0.93	0.97
NB	0.937	0.93	0.886	0.93	0.911	0.926	0.945	0.924
RF	**0.97**	0.96	**0.97**	0.95	**0.97**	0.95	**0.96**	0.96

4.4 Statistical Analysis of Results

The statistical analysis method that was used in this research is the Paired t-test and the Wilcoxon Signed-Rank Test.

4.4.1 The Paired T-Test

This is a statistical test used to compare the means of two related groups. It is usually used when there is a significant difference in the average results of two conditions or treatments applied to the same subjects. The paired t-test assumes that the differences between the paired observations are normally distributed. If the test yields a p-value less than 0.05, it considers the difference to be statistically significant, meaning that there is likely a real difference between the two groups.

4.4.2 The Wilcoxon Signed-Rank Test

This is a non-parametric test, meaning it does not assume a normal distribution for the data. It is used to compare two related groups when the paired t-test assumptions cannot be met, such as when the data is not normally distributed. Instead of comparing means, the Wilcoxon test evaluates the ranks of differences between paired observations and like the t-test, if the p-value is below 0.05, it suggests a significant difference between the two conditions.

4.4.3 The Result of the Analysis of Statistical Test

Paired Tests Results (WEKA vs Python on 70% Split)

Paired t-test Results

1. Accuracy: t = -1.40, p = 0.23
2. Recall: t = -0.90, p = 0.42
3. F1-Score: t = -0.77, p = 0.49
4. Precision: t = -1.01, p = 0.37

Wilcoxon Test Results (Non-parametric)
1. Accuracy: W = 2.0, p = 0.19
2. Recall: W = 5.0, p = 0.63
3. F1-Score: W = 5.0, p = 0.63
4. Precision: W = 3.0, p = 0.31

Results Interpretation

Both tests show no statistically significant difference between WEKA and Python environments for accuracy, recall, f1-score, and precision under the 70% split. All p-values were above 0.05, indicating that the performance differences between the environments are not significant.

Paired Tests Results (WEKA vs Python on 10-fold Cross-Validation)

Paired t-test Results
1. Accuracy: t = 0.81, p = 0.46
2. Recall: t = 1.76, p = 0.15
3. F1-Score: t = 1.55, p = 0.20
4. Precision: t = 0.76, p = 0.49

Wilcoxon Test Results (Non-parametric)
1. Accuracy: W = 6.0, p = 0.81
2. Recall: W = 2.0, p = 0.19
3. F1-Score: W = 3.0, p = 0.31
4. Precision: W = 4.0, p = 0.44

Results Interpretation

Both tests (t-test and Wilcoxon) show no significant difference between WEKA and Python environments for accuracy, recall, f1-score, and precision at the typical significance level ($p < 0.05$). The p-values are all above 0.05, suggesting that the performance variations between the environments for these classifiers are not statistically significant under 10-fold cross-validation.

4.5 Challenges in Achieving Generalizability

4.5.1 4.5.1 Class Imbalance

One of the key challenges in building ML models for breast cancer detection is the class imbalance, where the number of samples in one class significantly outweighs the number in the other class. This imbalance can lead to a bias in the model, causing it to favour the majority class. For example, a model might achieve high accuracy simply by predicting the majority class in most cases, while failing to correctly identify critical instances from the minority class. This study addresses class imbalance by making use of more than one metrics such as recall and precision to validate the performance of the classifiers. A high recall for malignant cases is particularly important in medical applications to minimize false negatives, which represent undetected cancer cases. Additionally, using metrics like the F1-score or the Area Under the Precision-Recall Curve (AUC-PR) instead of accuracy alone can provide a more balanced evaluation of model performance.

4.5.2 Overfitting

Another challenge in achieving generalizability is overfitting, where the model learns patterns specific to the training data, including noise, at the expense of generalizing to unseen data. This was particularly relevant in the case of the MLP model which showed strong performance in WEKA but inconsistent results in Python Jupyter Notebook. Overfitting often manifests as high accuracy on the training set but reduced performance on validation or test datasets. Overfitting can be aggravated by small dataset sizes or high model complexity. Techniques such

as cross-validation and regularization can help reduce the risk of overfitting. Additionally, increasing the dataset size or using data augmentation methods could further improve generalizability.

The choice of software environment, can also influence generalizability. Differences in algorithm implementations, preprocessing methods, validation and default parameters settings can lead to varying performance across environments. Addressing class imbalance and overfitting in a consistent and methodical way across environments.

5 Conclusion

The early detection of breast cancer stands as a leading factor that has significantly increased the survival rate in patients. The successful integration of ICT into the field of medical science has heralded the arrival of innovative technologies such as ML, deep learning, and AI. These technologies have unequivocally demonstrated their ability to provide faster and more efficient methods for detecting and predicting breast cancer, ultimately resulting in a marked increase in the survival rate of patients. This research aimed to investigate the generalizability of SVM models for breast cancer detection by comparing their performance with other classifiers such as NB, RF, MLP, and DT under various programming environments and data-splitting techniques. It was discovered that SVM performed the best under the 10-fold cross-validation and percentage split techniques in WEKA with accuracy values of 0.981 and 0.977 respectively. In the Jupyter Notebook, SVM had the highest accuracy value of 0.99 in the 70% split, but in 10-fold cross-validation, RF outperformed all other algorithms with an accuracy value of 0.97. It is worth noting that the SVM was not too far off, as it had an accuracy value of 0.96. Hence, SVM is recommended to be leveraged as a built-in algorithm for medical applications, which would be helpful to medical practitioners or clinicians for the early detection of breast cancer. Future research may explore the impact of additional factors like dataset size, class imbalance, and feature engineering on the generalizability of the models. Additionally, the research could be extended to incorporate more cutting-edge deep learning architectures for a more comprehensive evaluation of model performance in breast cancer detection. This research utilizes stronger language to emphasize the importance of the research and the potential impact of the research findings.

References

Akbuğday, B. (2019). Classification of breast cancer data using machine learning algorithms. 2019 Medical Technologies Congress (TIPTEKNO). https://doi.org/10.1109/tiptekno.2019.8895222

Chaurasiya, S., & Rajak, R. (2022). Comparative analysis of machine learning algorithms in breast cancer classification. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-1772158/v1

Dinesh, P., Vickram, A. S., & Kalyanasundaram, P. (2024). Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest and decision tree to measure accuracy. AIP Conference Proceedings. https://doi.org/10.1063/5.0203746

Elsadig, M. A., Altigani, A., & Elshoush, H. T. (2023). Breast cancer detection using machine learning approaches: a comparative study. International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, 13(1), 736. https://doi.org/10.11591/ijece.v13i1.pp736-745

Fatima, N., Liu, L., Sha, H., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. IEEE Access, 8, 150360–150376. https://doi.org/10.1109/access.2020.3016715

Guleria, K., Sharma, A., Lilhore, U. K., & Prasad, D. (2020). Breast cancer prediction and classification using supervised learning techniques. Journal of Computational and Theoretical Nanoscience, 17(6), 2519–2522. https://doi.org/10.1166/jctn.2020.8924

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-09954-8

Hoque, N. R., Das, N. S., Hoque, N. M., & Hoque, N. M. (2024). Breast Cancer Classification using XGBoost. World Journal of Advanced Research and Reviews, 21(2), 1985–1994. https://doi.org/10.30574/wjarr.2024.21.2.0625

Ibeni, W. N. L. W. H., Salikon, M. Z. M., Mustapha, A., Daud, S. A., & Salleh, M. N. M. (2019). Comparative analysis on Bayesian classification for breast cancer problem. Bulletin of Electrical Engineering and Informatics, 8(4). https://doi.org/10.11591/eei.v8i4.1628

Islam, T., Sheakh, M. A., Tahosin, M. S., Hena, M. H., Akash, S., Jardan, Y. a. B., FentahunWondmie, G., Nafidi, H., & Bourhia, M. (2024). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-57740-5

Keleş, M. K. (2019). Breast Cancer Prediction and detection Using Data Mining Classification Algorithms: A Comparative study. Tehnicki Vjesnik-technical Gazette, 26(1). https://doi.org/10.17559/tv-20180417102943

Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In Communications in computer and information science (pp. 108–117). https://doi.org/10.1007/978-981-15-7205-0_10

Mosayebi, A., Mojaradi, B., Naeini, A. B., & Hosseini, S. H. K. (2020). Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. PLOS ONE, 15(10), e0237658. https://doi.org/10.1371/journal.pone.0237658

Naji, M. A., Filali, S. E., Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., & Debauche, O. (2021). Machine learning Algorithms for breast cancer prediction and diagnosis. Procedia Computer Science, 191, 487–492. https://doi.org/10.1016/j.procs.2021.07.062

Shah, C., & Jivani, A. (2013). Comparison of data mining classification algorithms for breast cancer prediction. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). https://doi.org/10.1109/icccnt.2013.6726477

Strelcenia, E., & Prakoonwit, S. (2023). Effective feature engineering and Classification of breast cancer diagnosis: a Comparative study. BioMedInformatics, 3(3), 616–631. https://doi.org/10.3390/biomedinformatics3030042 Risk factors and preventions of breast cancer

Sun, Y., Zhao, Z., Zhang, Y., Fang, X., Lu, H., Zhu, Z., Shi, W., Jiang, J., Yao, P., & Zhu, H. (2017). Risk factors and preventions of breast cancer. International Journal of Biological Sciences, 13(11), 1387–1397. https://doi.org/10.7150/ijbs.21635

World Health Organization: WHO & World Health Organization: WHO. (2023, July 12). Breast cancer. https://www.who.int/news-room/fact-sheets/ detail/breast-cancer

Sign Language Recognition Using Residual Network Architectures for Alphabet And Diagraph Classification

[1*]Martins E. Irhebhude, [2]Adeola O. Kolawole and [2,3]Wali M. Zubair

[1,2]Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, Kaduna, Nigeria
[3]Department of Computer Science, College of Science Technology, Kaduna Polytechnic, Kaduna, Nigeria

email: [1*]mirhebhude@nda.edu.ng, [2]adeolakolawole@nda.edu.ng, [3]zubairwm@kadunapolytechnic.edu.ng

**Corresponding author*

Received: 08 October 2024 | Accepted: 11 December 2024 | Early access: 19 December 2024

Abstract - *Communication is crucial in human life, enabling the exchange of information through various methods beyond spoken language. Sign language translation is crucial for bridging communication gaps between hearing-impaired and hearing individuals, promoting effective interaction and understanding. This study presents a comprehensive model for identifying alphabet and digraph signs using feature extraction techniques from ResNet architectures, specifically ResNet18, ResNet50, and ResNet101. The system was designed to integrate both hand gestures and facial expressions, enhancing the accuracy of sign language recognition. Classification of sign language images into alphabet and diagraph categories was assessed using Support Vector Machine (SVM). The resulting classification accuracies were 61.7% for ResNet18, 64.5% for ResNet50, and 66.5% for ResNet101. The research results emphasize how deeper ResNet models are effective in improving recognition accuracy. This proposed model has significant implications for educational applications as it addresses attention-related challenges and aims to enhance student engagement in learning processes, thereby contributing to developing more inclusive educational environments.*

Keywords: Alphabet Sign, Diagraph Sign, ResNet, Hearing-Impaired, Sign Language Recognition, Support Vector Machine.

1 Introduction

Hand gesture recognition is used to create systems for exchanging information among individuals with disabilities or controlling devices.(Sahoo et al., 2021). According to Al-Hammadi et al. (2020), a significant application of hand gesture recognition is to facilitate the translation of sign language. A gesture is a physical movement involving the hands, arms, face, or body, aimed at conveying information or meaning (Irhebhude et al., 2023). Gesture recognition involves tracking human movements and interpreting them as meaningful commands with semantic significance. In the field of Human-Computer Interaction (HCI), two primary methodologies are used to interpret gestures: data glove techniques and vision-based methods (Ma et al., 2021).

In the data glove method, hand gesture data is collected using motion sensors, gloves, and trackers (Côté Allard et al., 2019). However, this approach is costly due to the need for hardware and can be cumbersome as it restricts the movement of signers. The vision-based method uses cameras and imaging sensors to gather data, unlike the other method (Haria et al., 2017). HCI is essential in information technology, involving computer-human interaction. Hand gestures are a nonverbal and innate way to interact with computers. Recognizing hand gestures is crucial in HCI, especially for those who are hearing impaired (Haria et al., 2017).

Hearing-impaired individuals use sign languages, which are visual-based natural languages, for communication. Given that most hearing individuals do not understand sign language, sign language translation (SLT) has become essential for facilitating communication between these two groups (Gupta & Singh, 2024). In recent years,

researchers have increasingly explored deep learning models for neural SLT to address this communication challenge.

Effective communication is essential for individuals to collaborate, express emotions and ideas, and contribute to societal advancement. Individuals with hearing impairments naturally develop sign language as a means of communication tailored to their needs (Sharma & Singh, 2020). Sign language, as a primal and innate form of communication, predates the early stages of human evolution. The development of sign language aligns with early historical theories, emerging even prior to spoken languages. Throughout history, sign language has evolved and seamlessly integrated into everyday communication, becoming an essential component of human interaction (Olabanji & Ponnle, 2021).

Sign language recognition systems benefit from an understanding of the hearing-impaired culture. Hearing-impaired culture encompasses the shared experiences, values, and traditions of a community where sign language serves as the primary mode of communication (Wen et al., 2021). This cultural awareness can lead to more accurate, reliable, and inclusive speech recognition (SLR) systems that promote accessibility and empowerment for individuals and communities with hearing impairment. The Nigerian hearing-impaired community was first introduced to American Sign Language (ASL) in the 1960s. However, according to Asonye et al. (2018), The indigenous Nigerian Sign Language (NSL) has historically been marginalized and misrepresented. NSL, which developed organically within the Nigerian hearing-impaired community, remains the primary means of communication for many Nigerians with speech and hearing difficulties.

Researchers have documented various regional varieties of NSL. For instance, Morgan (2002) studied Hausa Sign Language, utilized by hearing-impaired communities in northern Nigeria. Bura Sign Language, another regional sign language, has also been identified and analyzed in academic literature (Blench et al., 2006). The regional sign languages have a similar linguistic basis but show different vocabulary and grammar based on local spoken languages and cultural contexts. The preference for ASL over NSL has restricted the ability of hearing-impaired Nigerians to receive education, employment, and social services in their own language. Efforts to promote and preserve NSL as a vital component of the culture and identity of hearing-impaired Nigerians are crucial for ensuring their linguistic rights and inclusion in all aspects of society (Asonye et al., 2018).

Sign language is a visual-gestural mode of communication used by the hearing-impaired community. However, it faces significant challenges in facilitating effective communication between signers and non-signers. Recent advancements in the fields of deep learning and computer vision have been explored as potential solutions to address these challenges. However, as identified by Bragg et al. (2019), current research often excludes input from hearing-impaired individuals, who have firsthand experience with the challenges of sign language recognition algorithms. Additionally, Bragg et al. (2019) noted that many of the datasets used to train sign language recognition algorithms do not accurately represent real-world scenarios. Irhebhude et al. (2023) in their work, extracted diagraph signs from a school for the hearing-impaired in Kaduna State, Nigeria, and developed diagraph sign language recognition using a Residual Network (ResNet18) as a feature extractor and Support Vector Machine (SVM) as the classifier. The study did not, however, include recognition of alphabet signs. To overcome these limitations, the paper suggests using a real-world locally obtained dataset of the 26 English alphabets and some selected 16 diagraph signs, and applying ResNet18, ResNet50, and ResNet101.

The remainder of the paper is as follows: section II describes the Residual network models. Related studies are reported in section III. The methodology is discussed in section IV. Section V discusses the experimental results. The summary of findings, conclusion, and recommendations, are presented in section VI.

2 Residual Network

Convolutional Neural Networks (CNNs) consist of a wide range of architectures, such as the Residual Network family, which show variations within a foundational framework, primarily based on the number of layers and the total parameters (Hasanah et al., 2023). Residual Network (ResNet), is a deep convolutional neural network architecture introduced by He et al. (2016) in their 2015 paper "Deep Residual Learning for Image Recognition" (He et al., 2016). ResNet tackles the issue of vanishing gradients in deep neural networks by introducing "skip connections." These connections enable layers to learn residual functions based on the layer inputs, rather than learning unreferenced functions (He et al., 2016). The main advantage of ResNet is that it simplifies the optimization of extremely deep convolutional neural networks (Hasanah et al., 2023). ResNet models are named based on the depth of the network, such as ResNet-50 or ResNet-101, with the number denoting the number of layers in the model. Each level within the ResNet framework serves a distinct purpose that aids in the machine learning process and helps in extracting crucial features for classification assignments. It is widely known that

increasing the number of layers in a CNN architecture promotes deeper learning, which often leads to improved performance. However, this advantage comes with the drawback of longer training times due to the significant increase in parameters associated with deeper architectures (Hasanah et al., 2023).

2.1 ResNet-18

ResNet-18 is a foundational model within the ResNet family, known for its simplicity. It comprises 18 layers, with 17 convolutional layers and one fully connected layer. The architecture makes use of residual blocks, each containing two convolutional layers with 3 × 3 kernels, separated by batch normalization and ReLU activations. These residual blocks are characterized by a shortcut connection that bypasses the convolutional layers, enabling the network to learn residual mappings (He et al., 2016).

2.2 ResNet-50

ResNet-50 introduces a more complex architecture with 50 layers. It includes the use of Bottleneck blocks, which consist of three convolutional layers: a 1 × 1 convolutional layer for dimensionality reduction, a 3 × 3 convolutional layer, and another 1 × 1 convolutional layer for dimensionality expansion. The Bottleneck blocks help in reducing the computational complexity while maintaining high performance (He et al., 2016).

2.3 ResNet-101

ResNet-101 is an extension of the ResNet family, featuring a deeper network with 101 layers. It employs the same Bottleneck block structure as ResNet-50 but with an increased number of layers. The expanded architecture enables ResNet-101 to capture more intricate patterns and features, but it also demands greater computational resources and memory (He et al., 2016).

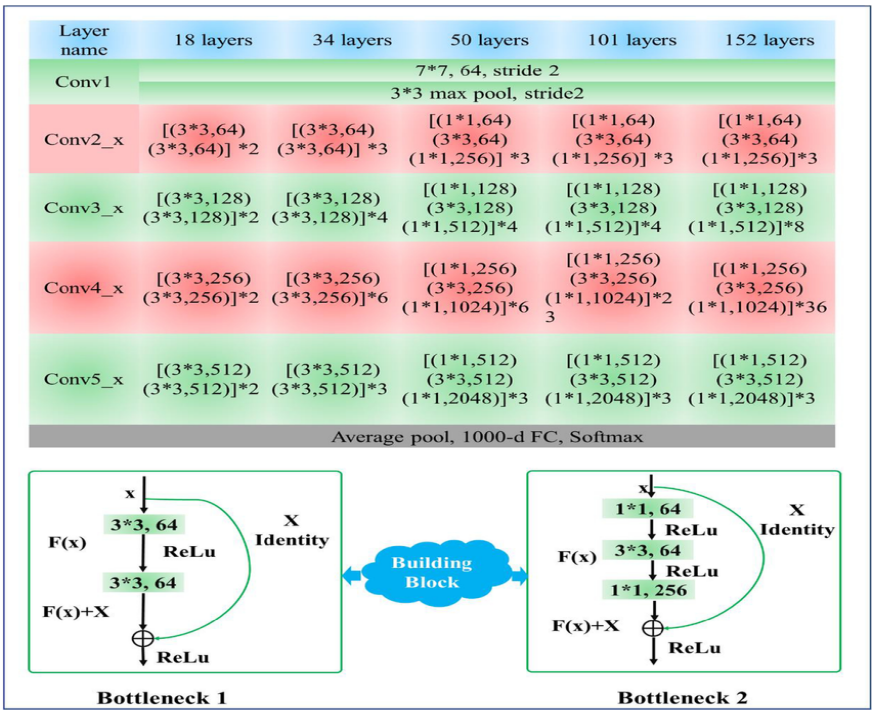


Figure 1: The structure of the ResNet: ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 (Lin et al., 2022).

The ResNet architecture, shown in Figure 1, consists of two main building blocks: bottleneck 1 and bottleneck 2. ResNet-18 and ResNet-34 use bottleneck 1, while bottleneck 2 is used in ResNet-50, ResNet-101, and ResNet-152. The number in the model name, such as 18 or 152, indicates the total number of layers in the model (He et al., 2016). In comparison to bottleneck 1, bottleneck 2 incorporates three convolutional layers of sizes 1 × 1, 3 × 3, and 1 × 1. The initial 1 × 1 convolutional layer functions to reduce the dimensionality of the input, making the 3 × 3 convolutional layer acts as a bottleneck with constrained input/output dimensions. The subsequent 1 × 1 convolutional layer then restores the dimensionality of the input to its original size (He et al., 2016).

The identity mapping in ResNet architectures is an important feature that helps mitigate the degradation problem. This issue arises when the training accuracy plateaus and then declines rapidly as the network becomes deeper. The identity mapping is achieved through skip connections, which allow the input of a layer to be added directly to its output, bypassing one or more layers. This allows the network to learn residual functions, which are easier to optimize than learning the original, unreferenced functions (He et al., 2016). The identity mapping in ResNet is defined as shown in equation 1:

$$Y = F(x, [\![wi]\!]) + X \quad (1)$$

where x is the input to the layer, $F(x, [\![wi]\!])$ is the residual function learned by the layer, and Y is the output of the layer. The addition of X to $F(x, [\![wi]\!])$ is the identity mapping, which ensures that the network can at least learn the identity mapping if the residual function is difficult to optimize.

ResNet18, ResNet50, and ResNet101 were chosen as pre-trained models and used as feature extracted to train an SVM classifier used for sign language recognition in this work.

3 Related Literature

This section discusses some of the related works on sign language recognition, the techniques, dataset, and results. Shi et al. (2022) proposed an innovative residual neural network architecture that integrates an enhanced residual module with a Bi-directional Long Short-Term Memory (BiLSTM) model to effectively classify 3D sign language gestures. The study evaluated this approach using challenging 3D sign language datasets from Chalearn and Sports-1M. The researchers devised a multi-path hybrid residual neural network architecture that combined improved residual modules for spatial feature extraction and BiLSTM for capturing temporal dynamics. The residual module incorporated motion excitation to enhance motion information captured and hierarchical-split blocks for extracting features across multiple scales. The hybrid neural network was trained end-to-end, assessed on benchmark datasets, and compared against state-of-the-art sign language recognition methods. Results showed that the model achieved a classification accuracy of 78.9% on the first dataset and 82.7% on the second dataset, surpassing existing algorithms and nearing human-level accuracy of 88.4%.

Sahoo et al. (2021) presented a hand gesture recognition system employing deep convolutional neural network (CNN) features integrated with machine learning techniques to develop a user-independent system capable of accurately recognizing hand gestures without requiring specific user training or calibration. The researchers' approach involved extracting deep CNN features from pre-trained models such as AlexNet and VGG-16, which were subsequently fed into a support vector machine (SVM) classifier. The authors investigated the utilization of CNN features from various layers independently and in combination to optimize hand gesture recognition accuracy. Evaluations conducted on a standardized American Sign Language (ASL) dataset demonstrated the system's effectiveness. Using features from the fully connected layers of AlexNet, the system achieved an accuracy of 92.8%. Combining features from multiple CNN layers further improved accuracy to 94.1%, surpassing benchmarks set by existing methodologies.

Jain et al. (2021) explored the use of SVM and CNN models to develop an effective system for recognizing ASL gestures. In the first phase, features from the dataset were extracted after applying various preprocessing techniques, using SVM with four different kernels i.e., ‘poly’, ‘linear’, ‘rbf’, and ‘sigmoid’. CNN with single and double layers was applied to the training dataset to train the model. Experimental results showed that the hybrid system using SVM and CNN achieved an accuracy of 98.58% in recognizing ASL gestures. The optimal filter size was found to be 8×8 for both single and double-layer CNN.

Venugopalan & Reghunadhan (2023) aimed to create an advanced sign language recognition system that can assist in improving communication and accessibility for hearing-impaired COVID-19 patients, who faced additional challenges in expressing their needs and concerns during the pandemic. The videos of hand gestures for the most common Indian sign language (ISL) words used for urgent communication by COVID-19-positive hearing-impaired patients were included in the proposed dataset. The proposed SLR utilized a deep learning model designed as a combination of the VGG-16 and bidirectional long short-term memory (BiLSTM) sequence network. The classification of gestures was done with a hybrid model of VGG-16 and BiLSTM networks and achieved an average accuracy of 83.36%. The model was also tested on another ISL word dataset as well as the Cambridge hand gesture dataset to further assess its performance and achieved promising accuracies of 97% and 99.34% respectively.

Chowdhury et al. (2017) created a novel method for converting Bengali Sign Language into readable text using a combination of Artificial Neural Network (ANN) and SVM techniques. The researchers collected a dataset of Bengali sign language gestures and video data were preprocessed. Microsoft Kinect was used to take the input, which is the hand sign performed in front of the camera. The captured hand sign was eventually recognized, after joint and wrist detection and by assessing the contours. The contour feature was extracted and presented to the SVM for the classification of the sign. The contour finding algorithm utilized the convex hull method, and the features extracted after detection were passed through the SVM for recognition. To validate the performance of the proposed model, a dataset of both male and female hand gesture images was utilized. Experimental results demonstrated 84.11% classification accuracy.

Jiang & Zhu (2019) developed an effective method for identifying and recognizing Chinese Sign Language (CSL) using wavelet entropy and SVM techniques. The researchers collected a dataset of CSL gestures, the sign language gesture data was preprocessed, wavelet entropy was used for feature extraction of the CSL gestures and the extracted wavelet entropy features were used as input to an SVM classifier. The experiment was implemented on 10-fold cross-validation and it yielded an overall accuracy of 85.69 ± 0.59%.

Sreemathy et al. (2023) introduced an effective system for continuous, word-level recognition of sign language through a blend of machine-learning approaches. The authors employed a proprietary image dataset comprising 80 static signs, encompassing a total of 676 images. The study proposed two distinct models: You Only Look Once version 4 (YOLOv4) and SVM integrated with MediaPipe. SVM utilized linear, polynomial, and Radial Basis Function (RBF) kernels. SVM with MediaPipe achieved an accuracy of 98.62%, while YOLOv4 achieved a higher accuracy of 98.8%, surpassing existing state-of-the-art methods in the field.

Al-Hammadi et al. (2020) developed an effective method for recognizing sign language gestures using deep learning techniques while focusing on efficient hand gesture representation. Two separate instances of 3D Convolutional Neural Networks (3DCNNs) were employed to capture distinct features: one focusing on detailed hand shapes and the other on broader body configurations. Multi-Layer Perceptron's (MLPs) and auto encoders were utilized to integrate and globalize these local features, followed by classification using the SoftMax function. Additionally, to mitigate the training expenses associated with the 3DCNN module, the researchers explored domain adaptation techniques and conducted extensive experiments to refine knowledge transfer efficiency. The efficacy of the proposed approach was evaluated on the King Saud University Saudi Sign Language (KSU-SSL) datasets. Experiments were carried out in two scenarios; signer dependent mode and signer-independent mode, with the MLP achieving accuracies of 98.62% and 87.69% respectively and the auto encoder achieving accuracies of 98.75% and 84.89%.

Kothadiya et al. (2022) introduced a deep-learning model designed to detect and interpret gestures as words. Specifically, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, which are feedback-based learning architectures, were employed to recognize signs from individual frames of Indian Sign Language (ISL) videos. The proposed model explores four distinct sequential configurations combining LSTM and GRU layers, leveraging the custom dataset, IISL2020. Among these configurations, the model featuring a single LSTM layer followed by a GRU layer achieved 97% accuracy across 11 different signs.

Aksoy et al. (2021) employed deep learning and image processing techniques to detect Turkish Sign Language gestures. The authors curated a dataset consisting of 10,223 images covering 29 letters from the Turkish Sign Language alphabet. The images were enhanced using various image processing methods to optimize them for educational purposes. The study culminated in the classification of these images using a range of architectures including CapsNet, AlexNet, ResNet-50, DenseNet, VGG16, Xception, InceptionV3, NasNet, EfficientNet, HitNet, SqueezeNet, and a specially designed TSLNet for this research. Among these models, CapsNet and TSLNet demonstrated the highest accuracy rates, achieving 99.7% and 99.6%, respectively, in their classification performance.

Olabanji & Ponnle (2021) introduced a system for interpreting the native sign language of Nigeria. The methodology consists of three main stages: dataset generation, application of computer vision methodologies, and the development of a deep learning model. A multi-class CNN was specifically crafted to train and interpret the indigenous signs. The evaluation was conducted using a custom dataset containing 15,000 images of selected native words. The experimental results demonstrated a strong performance from the interpretation system, achieving an accuracy of 95.67%.

Liao et al. (2019) presented a multimodal dynamic sign language recognition method based on a deep 3-dimensional residual ConvNet and BiLSTM networks, which was named as BiLSTM-3D residual network (B3D ResNet). This approach comprised three primary components. Initially, it localized the hand object within video

frames to reduce the computational complexity of network operations. Subsequently, the B3D ResNet automatically extracted spatiotemporal features from video sequences, analyzing these features to establish intermediate scores corresponding to each action within the sequences. Finally, through video sequence classification, it accurately identified dynamic sign language expressions. Experimental validation was conducted on test datasets, including DEVISIGN_D and SLR_Dataset. Results demonstrated that the proposed approach achieved state-of-the-art recognition accuracy (89.8% on DEVISIGN_D and 86.9% on SLR_Dataset). Moreover, the B3D ResNet effectively recognized intricate hand gestures across extensive video sequences, achieving high accuracy in identifying 500 Chinese hand sign language vocabularies.

Gupta & Singh (2024) proposed an automated method for recognizing Indian sign language (ISL) gestures in English. Initially, the hand region was isolated using the Grasshopper optimization algorithm (GOA) based on a skin color model. The effectiveness of segmentation was evaluated using three approaches: GOA-based skin color detection algorithm (SCDA), particle swarm optimization-based SCDA (PSO-SCDA), and artificial bee colony-based SCDA (ABC-SCDA). Subsequently, a database was established containing gestures representing individual English alphabets. For gesture recognition, the system was trained using a template-based matching approach. The classification was performed using both SVM and convolutional neural network (CNN) techniques. The proposed recognition method achieved the highest accuracy of 97.85% with GOA-SCDA, compared to 89.29% and 93.96% with PSO-SCDA and ABC-SCDA, respectively. Additionally, CNN surpassed SVM in classification performance, achieving an accuracy of 99.2% and a precision of 81.8%.

The study in AkanshaTyagi (2022) developed an extraction technique that used the Fast Accelerated Segment Test (FAST), and Scale-Invariant Feature Transformation (SIFT). FAST and SIFT were hybridized and used to detect and extract features with CNN reserved classification. Results obtained showed that excellent recognition accuracies on four different datasets.

Irhebhude et al. (2023) devised a diagraph sign language recognition system by employing a ResNet18 as a feature extractor and SVM as the classifier. The researchers utilized a proprietary dataset comprising 796 images depicting students expressing 16 diagraph sounds, which include two and three-letter words conveyed through sign language. The system achieved an accuracy of 79.3%.

Chao et al. (2019) introduced a behavior recognition approach aimed at addressing sign language recognition challenges. Drawing inspiration from Multi-Fiber Networks, the authors proposed a Convolutional Block Attention Module CBAM-ResNet neural network that extends the ResNet architecture using 3D convolutions and integrates a convolutional block attention module. To enhance channel information fusion, the fifth layer incorporated the 3D-ResNet structure, leveraging the strengths of Multi-Fiber Networks while compensating for their limitations. The proposed method was compared with models incorporating convolutional block attention modules, Convolutional Long Short-Term Memory (ConvLSTM), optical flow, and other methodologies. Achieving an accuracy of 83.3% on the Chinese Sign Language Recognition Dataset without optical flow. The proposed approach demonstrated a performance improvement of approximately 9% over Multi-Fiber Networks.

As such, the main contribution of this work is to present a new alphabet and diagraph sign language recognition system by employing Residual Network and SVM. A previous work that devised a diagraph sign language recognition system by employing a Residual Network (Irhebhude et al., 2023) has already shown that the proposed model could be implemented within educational curricula to address issues stemming from attention deficits and enhance students' engagement in learning activities. This paper expands upon the aforementioned conference paper incorporating the 26 English alphabets into diagraph signs for better classification and improved performance. Additionally, three Residual network architectures (ResNet-18, ResNet-50, and ResNet-101) specifically designed for the robust and efficient classification of sign language are presented.

4 Proposed Methodology

This study employed ResNet for the feature extraction and SVM for the classification. Irhebhude et al. (2023) employed the ResNet architecture for feature extraction from the image dataset and classified the extracted features using SVM. The proposed model by Irhebhude et al. (2023) (shown in Figure 2) was adopted by evaluating three variants of the ResNet architectures (ResNet-18, ResNet-50, and ResNet-101).

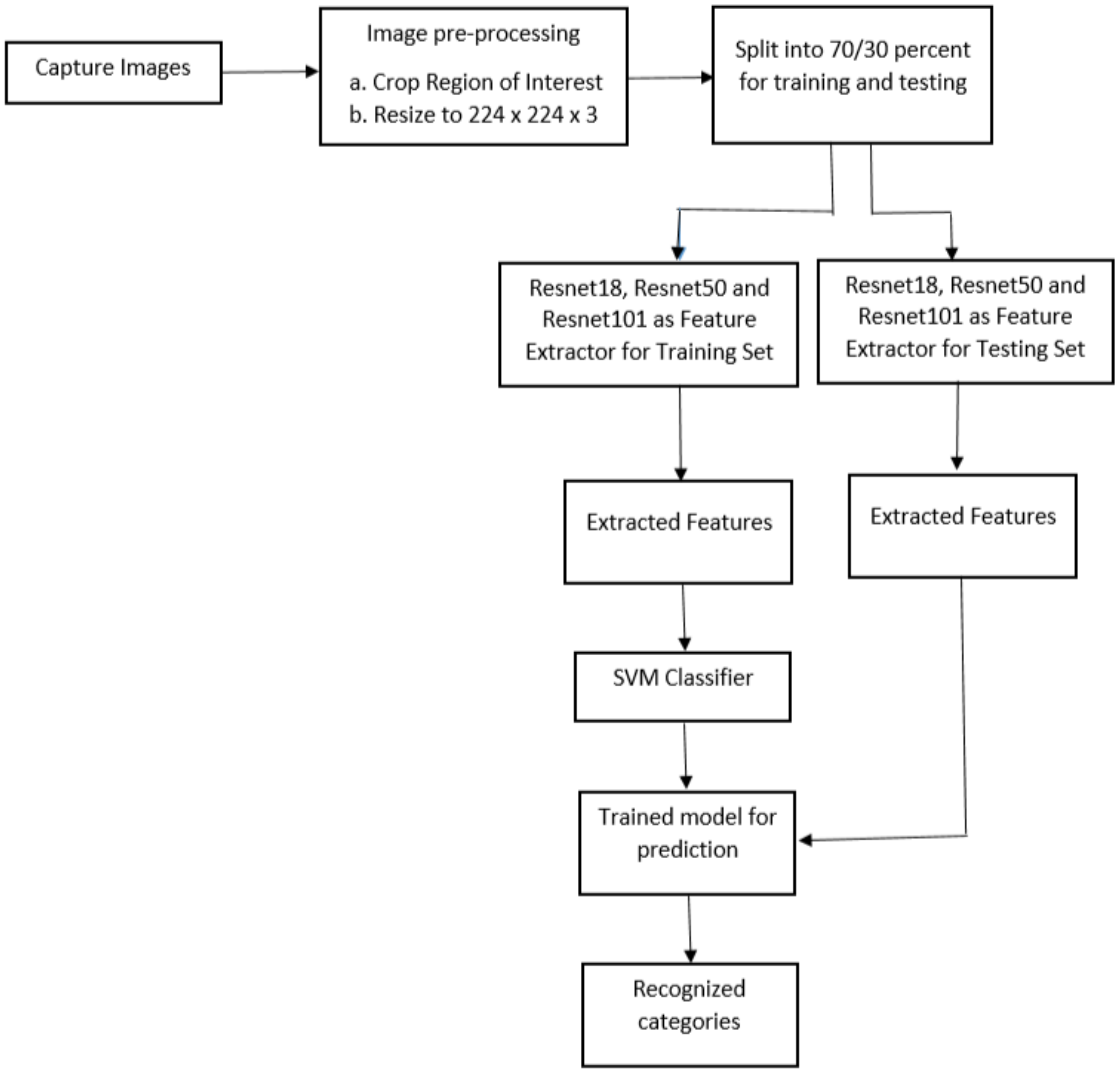


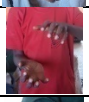
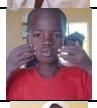
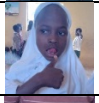
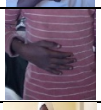
Figure 2: Proposed Methodology

The experiment used a dataset captured at LGEA Kagoro Road Primary School in Kaduna State, Nigeria. LGEA Kagoro Road Primary School was established in 1977, the school follows the basic 7 policies and primarily uses English and Hausa languages alongside British Sign Language (BSL). The school serves 1028 students aged 7 years and older, focusing on the English alphabet (a - z) and digraph sounds (like 'sh', 'ie', 'ch', 'ng', etc.). Despite demonstrations, students struggle to comprehend the material due to slow learning processes. Therefore, a vision-based technique was proposed (Figure 2) by Irhebhude et al. (2023) to interpret and demonstrate sign languages, aiming to benefit both hearing impaired and hearing individuals alike.

4.1 Image Capture

To compile the dataset used in the study, 26 English alphabets and 16 specific diagraph signs were photographed, each comprising a variety of images. Diagraphs are pairs of letters that work as a team to create a unique sound, different from the individual sounds of the letters (Daisie, 2023). The dataset comprises 2,106 images of students demonstrating the alphabet and diagraph expressed in sign language. These images were captured using a camera to capture both facial expressions and hand gestures of male and female students, with each image corresponding to a distinct alphabet and diagraph sign. Table 1 shows example images from each category, including a breakdown of the alphabet, diagraphs, and the number of images captured for each class, with the highest class having 51 images and the lowest class having 43 images, which is a fairly balanced distribution dataset. The variation in the number of images was as a result of wrong display of sign following ground truth information from the instructors.

Table 1: List of alphabets and diagraph sounds with image sample

Alphabet/ diagraphs	Sample	Number of Images
A		51
AI		51
AR		43
B		51
C		51
CH		51
D		50
E		51
EE		51
ER		49
F		51
G		51
H		50
I		50
IE		51
J		51
K		51
L		49
M		51
N		49
NG		50
O		51
OA		51
OI		49
OO		51
OOO		49
OR		51
OU		50
P		51
Q		51
R		51
S		50
SH		50
T		51
TH		48
U		51
UE		51
V		51
W		51
X		50
Y		46
Z		49

The primary objective is to identify and classify the sampled alphabet and diagraph signs. Figure 2 depicts the proposed methodology for the sign language recognition system showing all the steps involved while Table 1 describes the various alphabets/diagraphs signs dataset and the number of images captured in each category.

4.2 Image Pre-processing

Before data splitting and testing, it is crucial to preprocess the images. Initially, all captured images were of varying sizes and subsequently cropped and resized to standardized dimensions of 224 by 224 pixels, ensuring uniformity in their format. The dataset was then partitioned into a training set comprising 70% of the data and a testing set comprising the remaining 30%.

4.3 Feature Extraction

In this stage, the ResNet algorithm served as the feature extractor to derive sign language recognition features from the dataset. The ResNet used 18-layer, 50-layer, and 101-layer plain network architectures, detailed in Irhebhude et al. (2023), to extract deep learned features. The features were automatically extracted in the layer before the fully connected layer before the subsequent input to the SVM classifier for classification. Skip connections in ResNet improve deep neural network training and performance by maintaining gradient flow, reducing information loss, and increasing optimization and generalization, enabling training of networks to appropriate layers (Oyedotun et al. 2021; Zhang et al. 2020). ResNet as a top-performing feature extraction model due to its automatic, reliable, and versatile nature across multiple applications (Xu et al., 2022).

4.4 Classification

By learning key features for individual classes in the dataset, the SVM completes classification tasks for the alphabet and diagraph sign language. The alphabets and diagraphs were represented by a hand sign with facial expression which were recognized and the correct sign was identified by the models. SVM classifier is chosen for its efficiency, accuracy, robustness, and ability to utilize extracted features in image classification, particularly for large, high-dimensional datasets (Kashef, 2021). SVM efficiency and accuracy are influenced by exceptional feature extracted and optimal parameters, which can be improved through innovative feature selection methods (Wang et al., 2023).

5 Experimental Results and Discussions

The results of the analysis are presented in this section. The model was trained using 70% and 30% of the dataset for training and testing respectively. To experiment, the dataset consists of 26 alphabets and 16 classes of diagraph images. The words require both hand gestures and other parts of the face as shown in validation results shown in Figures 9-11. The classification model attained an accuracy of 61.7% for ResNet18, 64.5% accuracy for ResNet50, and 66.5% accuracy for ResNet101. The confusion matrix and Area Under Curve (AUC) of the Receiver Operating Curve (ROC) were used in evaluating the performance of the model. The hyperparameter tuning of the training is shown in Table 2.

Table 2: SVM Model Hyperparameters

Parameters	Value
Kernel Function	Linear
Box Constraint Level	1
Kernel Scale Mode	Auto
Multiclass Coding	One-vs-one
Standardise Data	Yes

The ROC curves depicted in Figures 3 to 5 illustrate the performance of the various models. Notably, ResNet50 exhibited the highest Area Under the Curve (AUC) with an impressive overall performance of 99.9%. Following closely, ResNet18 achieved an AUC of 97.4%. In contrast, ResNet101 showed the lowest AUC performance at

96.4%. These findings underscore ResNet50's exceptional classification capability, demonstrating its effectiveness in distinguishing between the various alphabets and diagraph sounds compared to the other models evaluated.

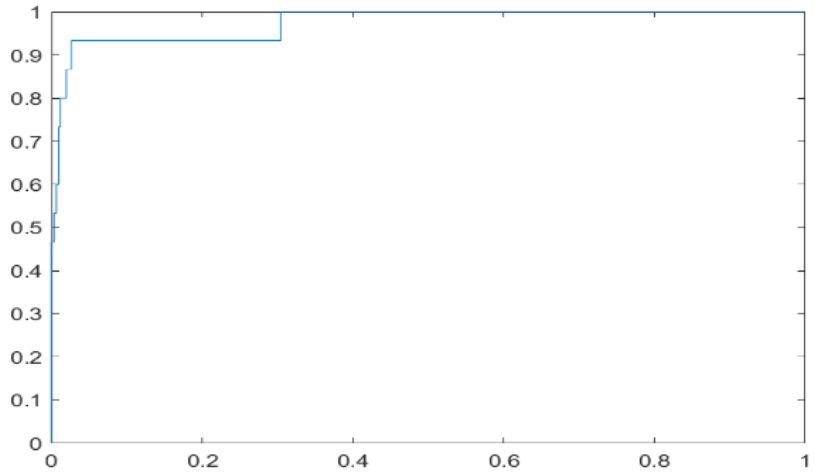


Figure 3: ROC Showing Classification Performance of ResNet18

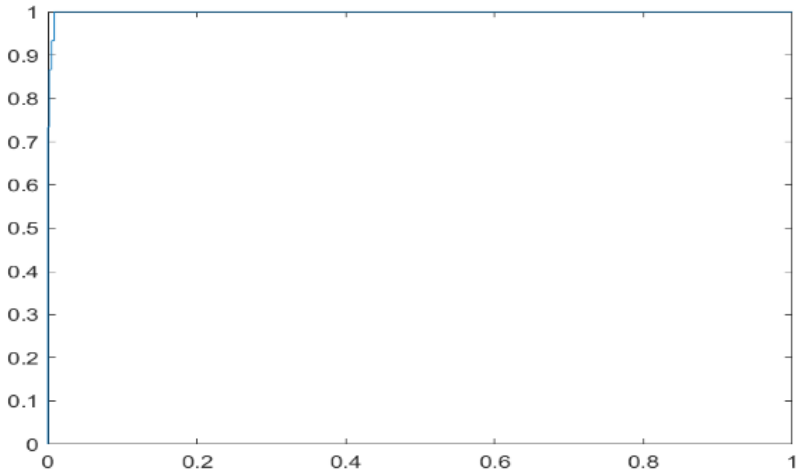


Figure 4: ROC Showing Classification Performance of ResNet50

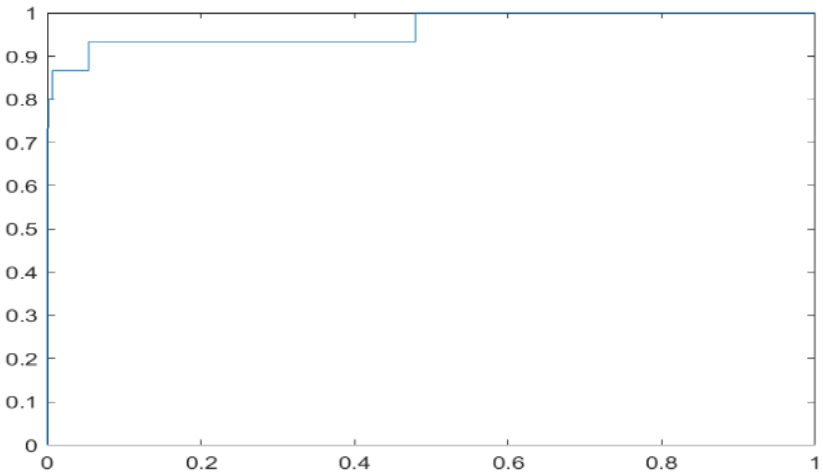


Figure 5: ROC Showing Classification Performance of ResNet101

The evaluation results, as presented in confusion matrices (Figures 6-8), provide insights into the performance of the models across different classes. A total of 632 sample images, comprising 30% of the entire dataset of alphabet

and diagraph images, were used for testing. This sample included 26 single-word alphabet classes, 15 two-word diagraph classes, and one three-word diagraph class, each contributing 30% of the test images.

Among the models evaluated, ResNet101 demonstrated the highest True Positive Rate (TPR) at 66.5% across all classes, indicating its capability to correctly identify positive instances. Conversely, it recorded a False Negative Rate (FNR) of 33.5%, indicating instances where positive instances were incorrectly classified as negative. In contrast, ResNet18 exhibited the lowest TPR of 61.7% and a corresponding FNR of 38.3% across all classes, suggesting its comparatively lower performance in correctly identifying positive instances.

Lastly, ResNet50 achieved a TPR of 64.5% and an FNR of 35.5%. These results collectively highlight ResNet101's superior performance in terms of correctly identifying positive instances, with a lower false negative rate compared to ResNet50.

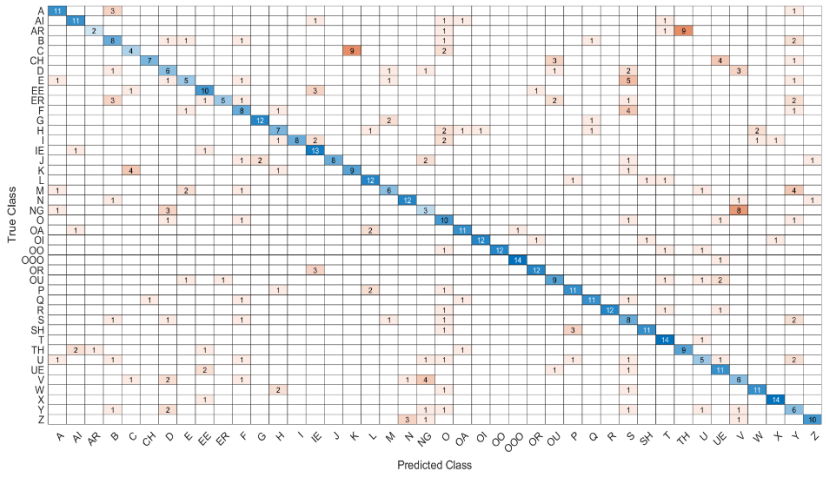


Figure 6: Confusion Matrix Showing Model Evaluation for ResNet18

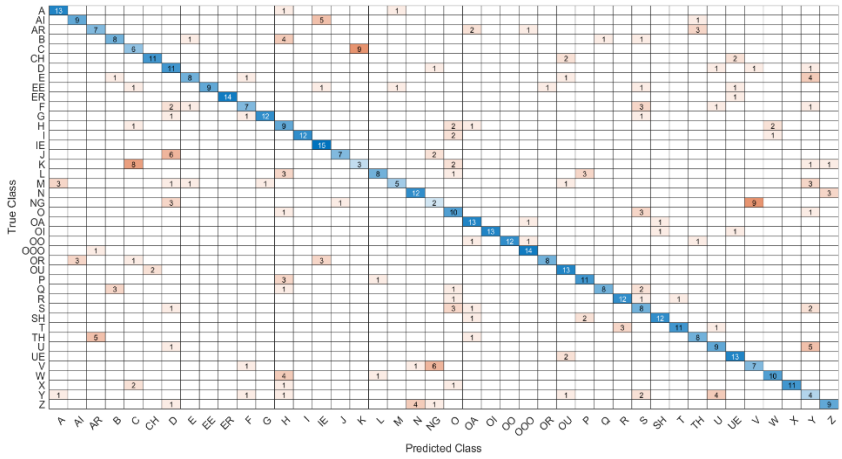


Figure 7: Confusion Matrix Showing Model Evaluation for ResNet50

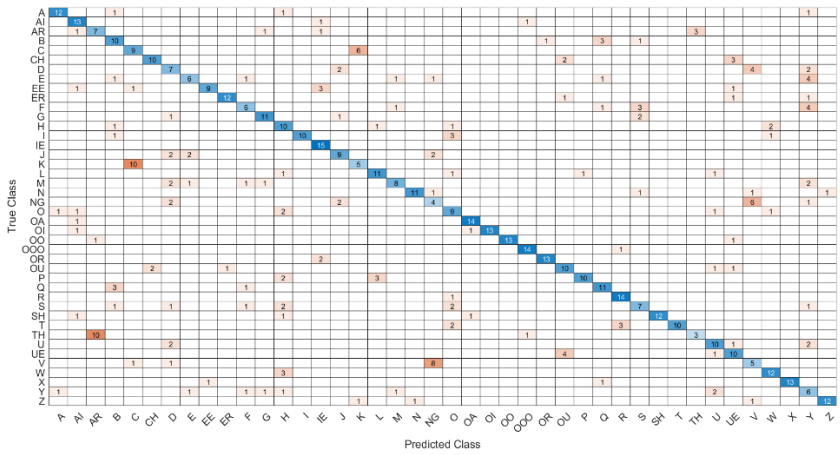


Figure 8: Confusion Matrix Showing Model Evaluation for ResNet101

These experiments demonstrated that the model achieved strong performance due to the quality of the input data and the effectiveness of the training process. The results further validate the proposed model's accuracy in interpreting alphabet and diagraph sign language, as evidenced by the findings presented in Figures 9-11. Results shows that few test images were wrongly predicted. The reason for this is the limited number of training data to enable the classifer learn to classifier into 42 groups.

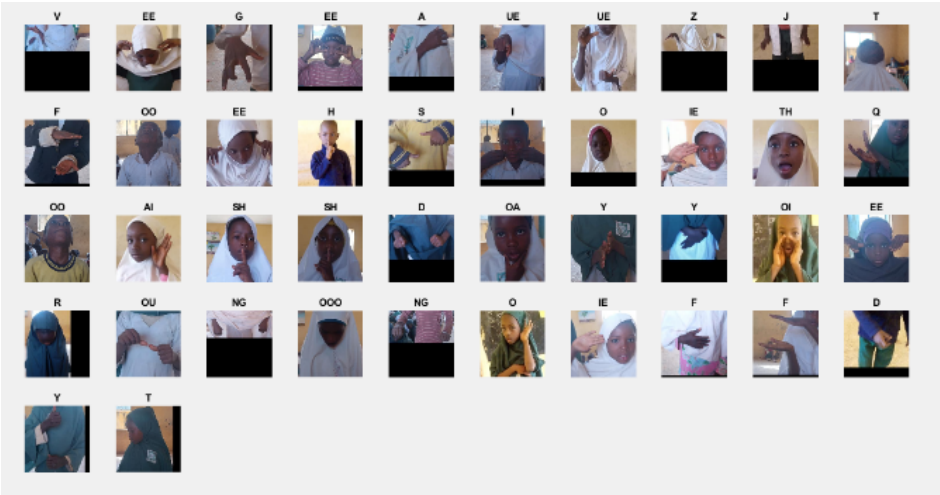


Figure 9: Validation Results for ResNet18

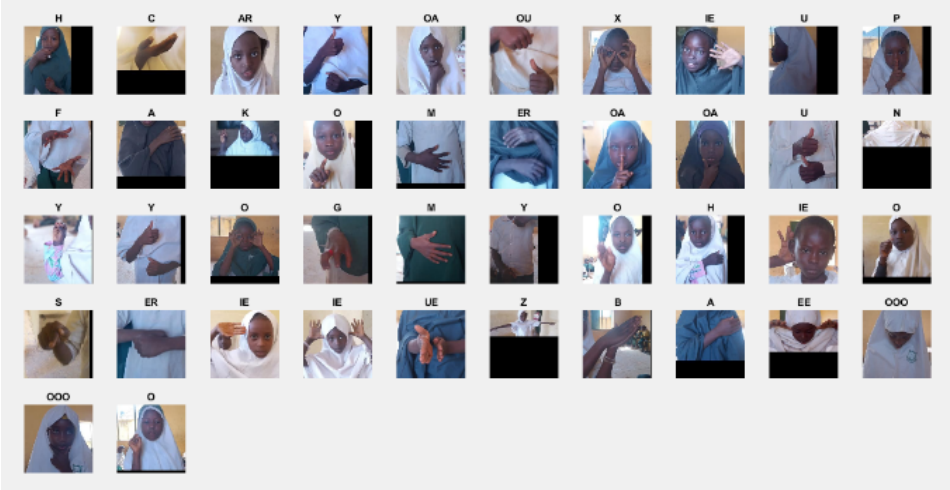


Figure 10: Validation Results for ResNet50

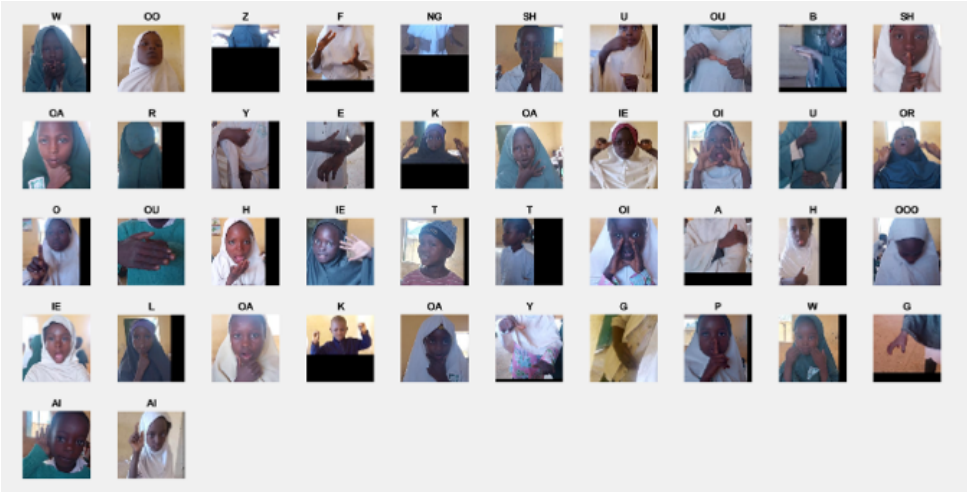


Figure 11: Validation Results for ResNet101

The results obtained from the experiments conducted on the captured dataset demonstrated the effectiveness of ResNet model. The ResNet101 model achieved the highest accuracy of 66.5% on the captured dataset compared to ResNet 18 and ResNet50 that gave 61.7% and 64.5% recognition accuracies respectively. The results indicate that the ResNet model can classify sign language images that include emotional cues.

The proposed model achieved the highest accuracy of 66.5% with ResNet101 on the self-captured dataset, highlighting the recognizing ability of ResNet101 with a higher number of layers. This result the higher layer model is particularly impressive when compared to the performance of the lower layer models. However, in terms of the performances of the classifier, ResNet50 performed best with an AUC of 99.9% when compared with the other models which gave 97.4% and 96.4% for ResNet18 and ResNet101 respectively. This difference in performance underscored the efficiency of the layer's depth.

The ResNet model's ability to achieve good accuracy on the dataset and excellent classification ability shows the robustness and adaptability of the model. This success suggests that the model can effectively learn and generalize the unique sign language for alphabet and diagraph signs, making it a more reliable choice for applications involving signs that incorporate gestures.

Olabanji & Ponnle (2021) achieved an accuracy of 95.67% on the Nigerian native sign language of the Nigeria classification using CNN. However, the study used a dataset for 15 selected words in Nigeria. The data collection procedure was not properly discussed for easy replication. The collected data only contains the hand sign avoiding the regions of the body that capture gesture information. Hence the need for this study. From the results obtained as shown from the confusion matrix of ResNet101 in Figure 8, the diagonal blue shows the true positive (TP) with diagraph IE recording highest true positive of 15 and TH recording the lowest TP of 3, while the highest false positive (FP) of 6 was recorded for alphabet C and NG diagraph. The false negative (FN) of 10 for the alphabet K and TH diagraph. Increasing layers impacted the performance of the experiments, with ResNet101 recording the highest recognition accuracy of 66.5%. This result validated what was obtained in the earlier study by Irhebhude et al. (2023).

6 Conclusion

In conclusion, this study introduced a model for recognizing alphabet and diagraph sign language, leveraging feature extraction from ResNet18, ResNet50, and ResNet101 algorithms. The system integrates hand gestures and facial movements for accurate sign language recognition. Classification into various alphabets and diagraph categories was achieved using SVM, yielding accuracies of 61.7% for ResNet18, 64.5% for ResNet50, and 66.5% for ResNet101 models, as evaluated on self-captured sign language images. This proposed model holds potential for integration into educational modules, addressing attention-related challenges and enhancing student engagement in learning processes. Given that sign language encompasses both spatial and temporal elements, with postures and gestures changing over time, integrating temporal data from video sequences may enhance recognition accuracy. It would be beneficial to capture and understand the temporal context of sign language using techniques like 3D convolutional neural networks (CNNs) or recurrent neural networks (RNNs).

References

Akansha Tyagi, S. B. (2022). Hybrid FiST_CNN Approach for Feature Extraction for Vision-Based Indian Sign Language Recognition. *The International Arab Journal of Information Technology (IAJIT)*, *19*(03), 403-411. https://doi.org/10.34028/iajit/19/3/15

Aksoy, B., Salman, O. K. M., & Ekrem, Ö. (2021). Detection of Turkish Sign Language Using Deep Learning and Image Processing Methods. *Applied Artificial Intelligence*, *35*(12), 952–981. https://doi.org/10.1080/08839514.2021.1982184

Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M., Alrayes, T., Mathkour, H., & Mekhtiche, M. (2020). Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation. *IEEE Access*, *8*, 192527–192542. https://doi.org/10.1109/ACCESS.2020.3032140

Asonye, E., Emma-Asonye, E., & Edward, M. (2018). Deaf in Nigeria: A Preliminary Survey of Isolated Deaf Communities. *SAGE Open*, *8*, 215824401878653. https://doi.org/10.1177/2158244018786538

Blench, R., Warren, A., & Dendo, M. (2006). *An unreported African Sign Language for the Deaf among the Bura in Northeast Nigeria*.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudrealt, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., & Morris, M. (2019). *Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective*.

Chao, H., Fenhua, W., & Ran, Z. (2019). Sign Language Recognition Based on CBAM-ResNet. *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, 1–6. https://doi.org/10.1145/3358331.3358379

Chowdhury, A. R., Biswas, A., Hasan, S., Rahman, T. M., & Uddin, J. (2017). Bengali Sign language to text conversion using artificial neural network and support vector machine. *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, 1–4. https://doi.org/10.1109/EICT.2017.8275248

Côté Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., Laviolette, F., & Gosselin, B. (2019). Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *PP*. https://doi.org/10.1109/TNSRE.2019.2896269

Daisie. (2023, June 21). *Diagraphs Explained: Comprehensive Phonics Guide*. Daisie Blog. https://blog.daisie.com/what-is-a-diagraph-a-comprehensive-guide-to-understanding-and-using-diagraphs-in-phonics/

Gupta, A. K., & Singh, S. (2024). Hand Gesture Recognition System Based on Indian Sign Language Using SVM and CNN. *International Journal of Image and Graphics*, 2650008. https://doi.org/10.1142/S0219467826500087

Haria, A., Subramanian, A., Asokkumar, N., Poddar, S., & Nayak, J. (2017). Hand Gesture Recognition for Human Computer Interaction. *Procedia Computer Science*, *115*, 367–374. https://doi.org/10.1016/j.procs.2017.09.092

Hasanah, S. A., Pravitasari, A. A., Abdullah, A. S., Yulita, I. N., & Asnawi, M. H. (2023). A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image. *Applied Sciences*, *13*(24), Article 24. https://doi.org/10.3390/app132413111

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. doi:10.1109/cvpr.2016.90

Irhebhude, M. E., Kolawole, A.O., & Abubakar, H. (2023). DIAGRAPH SIGN LANGUAGE RECOGNITION USING RESIDUAL NETWORK AND SUPPORT VECTOR MACHINE. *International Conference on Communication and E-Systems for Economic Stability | CeSES' 2023*. Retrieved May 9, 2024

Irhebhude, M., Kolawole, A., & Goshit, N. (2023). *Perspective on Dark-Skinned Emotion Recognition Using Deep-Learned and Handcrafted Feature Techniques* (pp. 1–24). https://doi.org/10.5772/intechopen.109739

Jain, V., Jain, A., Chauhan, A., Kotla, S. S., & Gautam, A. (2021). American Sign Language recognition using Support Vector Machine and Convolutional Neural Network. *International Journal of Information Technology*, *13*, 1193–1200. https://doi.org/10.1007/s41870-021-00617-x

Jiang, X., & Zhu, Z. (2019). *Chinese Sign Language Identification via Wavelet Entropy and Support Vector Machine*. 726–736. https://doi.org/10.1007/978-3-030-35231-8_53

Kashef, R. (2021). A boosted SVM classifier trained by incremental learning and decremental unlearning approach. *Expert Systems with Applications*, *167*, 114154. https://doi.org/10.1016/j.eswa.2020.114154

Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A.-B., & Corchado, J. M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. *Electronics*, *11*(11), 1780. https://doi.org/10.3390/electronics11111780

Liao, Y., Xiong, P., Min, W., Min, W., & Lu, J. (2019). Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks. *IEEE Access*, *7*, 38044–38054. https://doi.org/10.1109/ACCESS.2019.2904749

Lin, K., Zhao, Y., Gao, X., Zhang, M., Zhao, C., Peng, L., Zhang, Q., & Zhou, T. (2022). Applying a deep residual network coupling with transfer learning for recyclable waste sorting. *Environmental Science and Pollution Research*, *29*(60), 91081–91095. https://doi.org/10.1007/s11356-022-22167-w

Ma, R., Zhang, Z., & Chen, E. (2021). Human Motion Gesture Recognition Based on Computer Vision. *Complexity*, *2021*, 1–11. https://doi.org/10.1155/2021/6679746

Morgan, R. Z. (2002). Maganar Hannu: Language of the Hands: A Descriptive Analysis of Hausa Sign Language (review). *Sign Language Studies*, *2*(3), 335–341. https://doi.org/10.1353/sls.2002.0011

Olabanji, A., & Ponnle, A. (2021). Development of A Computer Aided Real-Time Interpretation System for Indigenous Sign Language in Nigeria Using Convolutional Neural Network. *European Journal of Electrical Engineering and Computer Science*, *5*, 68–74. https://doi.org/10.24018/ejece.2021.5.3.332

Oyedotun, O. K., Ismaeil, K. A., & Aouada, D. (2021). Training very deep neural networks: Rethinking the role of skip connections. *Neurocomputing*, *441*, 105–117. https://doi.org/10.1016/j.neucom.2021.02.004

Sahoo, J., Ari, S., & Patra, S. (2021). *A user independent hand gesture recognition system using deep CNN feature fusion and machine learning technique* (pp. 189–207). https://doi.org/10.1016/B978-0-12-822133-4.00011-6

Sharma, S., & Singh, S. (2020). Vision-based sign language recognition system: A Comprehensive Review. *2020 International Conference on Inventive Computation Technologies (ICICT)*, 140–144. https://doi.org/10.1109/ICICT48043.2020.9112409

Shi, X., Jiao, X., Meng, C., & Bian, Z. (2022). 3D Sign language recognition based on multi-path hybrid residual neural network. *2022 14th International Conference on Machine Learning and Computing (ICMLC)*. https://doi.org/10.1145/3529836.3529943

Sreemathy, R., Turuk, M., Chaudhary, S., Lavate, K., Ushire, A., & Khurana, S. (2023). Continuous word level sign language recognition using an expert system based on machine learning. *International Journal of Cognitive Computing in Engineering*, *4*, 170–178. https://doi.org/10.1016/j.ijcce.2023.04.002

Venugopalan, A., & Reghunadhan, R. (2023). Applying Hybrid Deep Neural Network for the Recognition of Sign Language Words Used by the Deaf COVID-19 Patients. *Arabian Journal for Science and Engineering*, *48*(2), 1349–1362. https://doi.org/10.1007/s13369-022-06843-0

Wang, J., Wang, X., Li, X., & Yi, J. (2023). A Hybrid Particle Swarm Optimization Algorithm with Dynamic Adjustment of Inertia Weight Based on a New Feature Selection Method to Optimize SVM Parameters. *Entropy*, *25*(3), 531. https://doi.org/10.3390/e25030531

Wen, F., Zhang, Z., He, T., & Lee, C. (2021). AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nature Communications*, *12*(1), 5378. https://doi.org/10.1038/s41467-021-25637-w

Xu, Y., Yang, W., Wu, X., Wang, Y., & Zhang, J. (2022). ResNet Model Automatically Extracts and Identifies FT-NIR Features for Geographical Traceability of Polygonatum kingianum. *Foods*, *11*(22), 3568. https://doi.org/10.3390/foods11223568

Zhang, W., Quan, H., Gandhi, O., Rajagopal, R., Tan, C.-W., & Srinivasan, D. (2020). Improving Probabilistic Load Forecasting Using Quantile Regression NN with Skip Connections. *IEEE Transactions on Smart Grid*, *11*(6), 5442–5450. https://doi.org/10.1109/TSG.2020.2995777

A Visually Impaired Mobile Application for Currency Recognition using MobileNetV2 CNN Architecture

[1*]Abraham Eseoghene Evwiekpaefe and [2]Habila Isacha

Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, Kaduna.

email: [1*]aeevwiekpaefe@nda.edu.ng, [2]isachahabila@nda.edu.ng

**Corresponding author*

Received: 24 December 2024 | Accepted: 25 February 2025 | Early access: 10 March 2025

Abstract - *There are at least 2.2 billion people with visual impairment globally of which almost half of these cases could have been addressed or prevented. Visual impairment has both personal and economic impacts on individuals. It impacts negatively on the quality of life, especially among adults. Many visually impaired persons in our society today face a lot of challenges and one of these challenges is object recognition. The visually impaired persons need assistance so they can perform monetary transactions without being cheated. This work is aimed at developing a model for the recognition of Nigerian naira notes for visually impaired persons. The model was trained using the concept of transfer learning with a trainable layer built on the MobileNetV2 convolutional neural network architecture pre-trained model using python programming language on Spyder anaconda IDE. The model was saved and converted to TensorFlow lite format which was deployed into a mobile application coded in the java programming language in android studio. A total of 3615 image datasets were collected, including N5, N10, N20, N50, N100, N200, N500, and N1000 denominations and some random images of objects that constitute the non-currency class for the training of the model. The collected data was divided into 80% for training and 20% for testing. The model achieved an accuracy of 98%.*

Keywords: Currency, Naira, Visually Impaired, Transfer learning, Deep learning, MobileNetV2.

1 Introduction

There are at least 2.2 billion people with visual impairment globally of which almost half of these cases could have been addressed or prevented. Vision loss can affect people of all ages but majority of those with visual impairments are above 50 years of age. Visual impairment has both personal and economic impact on individuals. It negatively impacts the quality of life, especially among adults. In the case of older adults, visual impairment contributes to their social isolation, difficulty in walking and navigation, etc (WHO, 2021).

The Nigerian National Blindness and visual Impairment survey conducted between 2005 and 2007 in Nigeria, revealed that there are at least 1.13 million individuals aged at least 40 years who are currently blind while at least 2.7 million adults aged at least 40 years have moderate visual impairment. Therefore, Nigeria, which is the most populated African country with diverse ethnic and cultural background, faces a growing public health problem; blindness and visual impairments (Akano, 2017).

In our society today, the visually impaired persons face difficulties in identifying objects (Mallikarjuna et al., 2021). It is easy for people with a functional sight to recognize a currency, but this is not the case for the visually impaired. A visually impaired person is someone with either a partial or complete vision disorder. They face difficulties in their daily activities especially recognizing things they can't see (Samant et al., 2020). Currency transaction is an indispensable part of human civilization (Tasnim et al., 2021).

There are many assistive devices developed for the visually impaired recently with few solutions in place to help them recognize objects in their environments particularly indoors. Though research in object recognition and detection is gaining grounds, computer vision-based solutions seems to be promising because of their ease of accessibility. Visual impairment refers to vision loss that constitutes a significant limitation of visual capability resulting from disease, trauma, or a congenital or degenerative condition that cannot be corrected by conventional means, including refractive correction, medication, or surgery. Loss of vision has an extremely strong effect on an individual's ability to carry out certain activities for example, navigation, accessing information, and recognizing objects both animate and inanimate objects in his/her surroundings. With extremely weakened or non-existent of a sense of sight, visually impaired people have to depend on their memory and also rely on their senses such as hearing, touch, taste and smell in order to locate or identify objects in their surroundings. The amount of information that is received via human vision is much larger and its processing time takes less time as compared to the rest of the human senses, to depend on these other senses for perception can result in spending much time (Jafri et al., 2013).

Yadav et al. (2020) reported that people with visual impairments suffer from identifying currency values. Many visually impaired persons in our society today face a lot of challenges, one of these challenges is object recognition. Monetary transactions are very sensitive especially when it comes to trusting those handling it for one (Mallikarjuna et al., 2021). Technological advancements have brought about the increase in cases associated with counterfeit currencies around the world. In Nigeria, the issue of counterfeit currency is the biggest challenge faced in cash transactions (Ogbuju et al., 2020). According to Sarfraz et al. (2019) advancement in digital imaging technology which includes the ability to print highly coloured papers has made it possible to produce fake currency banknotes. The visually impaired persons need assistance so they can perform monetary transactions on their own or at least know the value of the currency that could be given to them. Another problem is that one may be given a counterfeit currency and may not be able to identify whether the currency is counterfeit or genuine which may lead to cheating those who accept these counterfeit currencies. This raises the need for currency recognition and detection system so the visually impaired can distinguish among the values of currencies. The proposed system will address the problem faced by the visually impaired in their ability to identify and recognize currency values in order to avoid them be cheated by a third party when given a currency. The visually impaired who cannot see nor identify objects properly can easily be cheated by simply giving them any currency note or even paper. Therefore, the proposed system should be able to identify eight (8) Naira currency denominations and voice out the value of the denominations in English language.

2 Literature Review

2.1 Convolutional Neural Network

A deep learning network design known as a convolutional neural network (CNN or ConvNet) learns directly from data and does away with the requirement for manual feature extraction. CNNs can recognize objects, faces, and scenes in photos by looking for patterns in the images. CNNs are mostly used in computer vision and object recognition applications such as self-driving cars, facial recognition, and systems for classifying and recognizing currencies. A CNN, like other neural networks, is made up of an input layer, an output layer, and numerous hidden layers in between as shown in Figure 1.

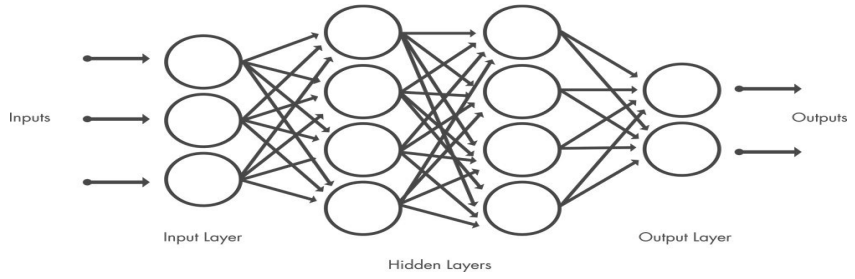


Figure 1: Typical CNN (Mathworks, 2021)

2.2 Transfer Learning

Is a concept that applies the known information or knowledge utilized or received from one operation to subsequent processes that are similar, which can drastically minimize the amount of data required and allow for a lightweight model design. E.g. MobileNet and RestNet (Zhu et al. 2021).

The workflow of the Transfer Learning Model is shown in Figure 2.

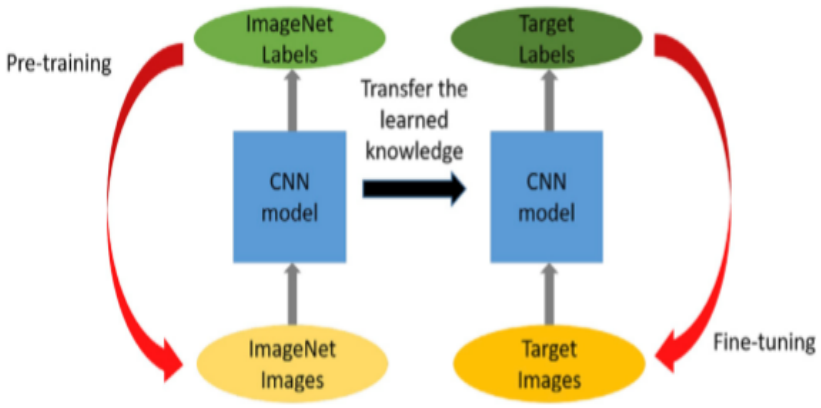


Figure 2: Transfer Learning workflow (Zhu et al. 2021)

In Figure 2, the transfer learning uses the pre-trained model based on the CNN architecture such as MobileNet or ResNet to obtain similar features from the data and train upon it. The transfer learning model gives room for fine-tuning the dataset in case the accuracy is low.

2.3 MobileNet CNN Architecture

The MobileNet model is based on depth wise separable convolutions which is a form of factorized convolutions which factorize a standard convolution into a depth wise convolution and a 1×1 convolution called a pointwise convolution. Depth wise convolution in MobileNets applies a single filter to each input channel. The depth-wise convolution's outputs are then combined using an 1x1 convolution by the pointwise convolution. A standard convolution both filters and puts together inputs into a new category of outputs in one stage. The depth wise separable convolution divides this into two layers, a different layer for filtering and another separate layer for combining. MobileNets are built on a simplified architecture that makes use of depth wise separable convolutions to build light weight. MobileNets are built primarily from depth wise separable convolutions and subsequently used in Inception models to reduce the computation in the first few layers. Down sampling is handled with stride convolution in the depth wise convolutions as well as in the first layer. A final average pooling reduces the spatial resolution to 1 before the fully connected layer. Counting depth wise and pointwise convolutions as separate layers, MobileNetV1 has 28 layers and MobileNetV2 has a total of 53 layers deep (Sandler et al., 2018).

Table 1: The MobileNetV2 architecture (Sandler et al., 2018)

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Table 1 shows the MobileNetV2 architecture, the parameter t is the expansion rate of the channels which uses a factor of 6, c is the number of input channels and n represents how often the block is repeated and s represents the first down sampling repetition of a block with a stride of 2. The architecture shows that input images have to be in the 224x224 dimension and in Red Green Blue (RGB) colour which is 3. As the training goes by, the architecture breaks down the dimension and for better high-level feature extraction (Sandler et al., 2018).

2.4 Review of Related Literature

2.4.1 Currency Detection for Visually Impaired Persons

According to Samant et al. (2020) it is easy for people who has functional sight to recognize a currency, but it is not the case for the visually disadvantaged. A visually impaired person is someone with either partial or complete sight disorder. They face difficulties in their daily activities especially recognizing things they can't see. They developed a system that is aimed at providing a cost-effective solution for visually impaired persons to be able to recognize currency using image processing techniques in form of an android application. The system works using voice commands and TensorFlow was used for the implementation. The system identifies and voices out the value of the currency.

According to Almu and Muhammad (2017) the manual method of identifying currency in Sokoto Metropolis is done manually by checking for certain features. This pose the challenge of differentiating between the original and fake currency. They developed a system using an image-based processing technique to help identify and recognize different currencies. The system was implemented using visual basic and Microsoft Access database management system. To test whether a currency is fake or not, the currency is fed into the system and the button start detection is clicked. The system returns a percentage of a match to a currency dataset. If the percentage of the match is high say 70 and above, the currency in question is considered genuine otherwise it Is considered as fake. The proposed future work includes the usage of two or more image processing algorithms to compare currencies for higher accuracy and the implementation on mobile devices and to identify bad or damaged currencies.

Suranya et al. (2020) implemented a currency counting system for visually impaired persons using SIFT algorithm along with ROI and OCR. The system scans a currency denomination and sums up the total then voices out the sum to the user. The system recognizes Indian currency. It is developed using the python packages and deployed into a hardware device which is assembled into a blind-stick with the camera on top of the sick for scanning the currency after which it is passed to the hardware device to echo the currency using a speaker. The hardware used was a raspberry pi with inbuilt Bluetooth and Wi-Fi. KNN algorithm was also used with 93.71% accuracy with good processing time but still has the limitation of differentiating the fake currency notes. This needs to be implemented in the system.

Ng et al. (2020) proposed an intelligent banknote recognition with assistive technology for visually impaired people using Hong Kong Dollar which is the main currency used in Hong Kong. Although the size of the banknotes differs but still it becomes difficult for the visually impaired to recognize them. Cheating can also be easier in case of monetary transactions involving the visually impaired. They applied a machine learning. The system has two parts: the first part is the pre-trained CNN model which was used for transfer learning which speeds up the time required for training. The system works by collecting input as a captured image, then image is pre-processed the feature extraction and the image passes through a binary classifier which determines whether the object is the Hong Kong banknote or not. If it is, then the note passes through a multi-class classifier which determines the denomination of the bank note. The application also checks the confidence level of the result of the classifier. If the confidence is not high enough say less than 70%, it requests the user to try again or else it gives voice and vibration feedbacks. The system was tested and has the accuracy of greater than 80% with the response time of less than 3 seconds. It was deployed using TensorFlow Lite Framework.

Islam et al. (2021) presented a new method for recognizing Bangladeshi currency using a Raspberry Pi 4 and the faster R-CNN model for the visually impaired. They built a smart blind glass that can recognize Bangladeshi currency and echo out the currency value in the blind man's ear. They also introduced a widely used model for object and face recognition which is the faster R-CNN approach. The faster R-CNN approach was used in training the dataset and has achieved a real-time accuracy of 97.8% which was considered high among the state-of-art research. They are with the opinion that visually impaired persons get cheated when it comes to monetary transactions. They used 4507 images as a dataset, these images are of different sizes and positions. 3571 images were used for training and 936 images for testing. The Pi camera captures the video immediately it senses the blind man holding money in his hand and the microprocessor processes the image and recognizes the amount then sends the sound of the amount to the speaker in the blind man's ear. The distance between the currency and the glass should be at last 60cm or less than that for proper focus. However, the blind man safety issues using voice recognition should be considered.

Tasnim et al. (2021) proposed an automated system for recognizing Bangladeshi currency using convolutional neural network for visually impaired persons. The datasets consist of 70, 000 bank notes of available Bangladeshi

currency. The system has an accuracy of 92% and gives the result with both audio and text outputs. For the implementation technique, for both training and testing, the Keras framework was applied with TensorFlow as the background. The CNN contains architecture which includes two sections which are automatic feature extraction method and classification. ConvNets contains a sequential architecture used to create the model. The pre-processed dataset is divided into 80% training and 20% testing. However, there is a fluctuation of accuracy due to the nature of the datasets. There is also a need to integrate the system into a cost effective and lightweight mobile application that can help blind people in daily transactions, they also recommended testing the dataset with other classifiers.

According to Gamage et al. (2020) the distortion of currencies over time has become a major challenge for recognizing currency banknotes especially to the visually impaired in Sinhala. In order to address this problem, they conducted research and implemented a system which comprises of three modules which include a speech recognition module, currency recognition module and text to speech module. Which aim to achieve better accuracy in all the three modules using deep learning. Speech recognition neural network model was built using TensorFlow platform and Keras library and deep learning neural networks were used for the development of currency recognition and text to speech modules. They further opined that though there are many speech recognition applications in English, applications which are built in native languages are rarely developed that gave them the reason to develop a speech recognition system in Sinhala language for the Sri Lankan currency to help the visually impaired. 9000 images were used for training of the model with the final validation accuracy of 87.80% and 3000 images were used for testing with accuracy of 95.90%. Jupyter notebook was used to train the model. Many voices should be added and a mobile application integrating the three modules need to be implemented for the visually impaired people to use the system.

Veeramsetty et al. (2020) developed a novel lightweight CNN model to recognize Indian currency notes and were deployed into a web and mobile applications. The proposed model was developed using Tensorflow which is improved by the selection of optimal hyperparameter value and was compared with some well CNN architectures using transfer learning. The created datasets of Indian currency to be used in the system. They developed a mobile application which was created using MobileNet CNN architecture which provides both prediction probability and audio outputs for the visually impaired. The work can be further extended by incorporating the identification mechanism for counterfeit currency notes. This can be done by extracting the face value of currency note and other features which cannot be incorporate in counterfeit notes.

Naing et al. (2019) proposed a Smart Blind Walking Stick Using Arduino which was based on sensors and microcontroller. The system helps the blind persons navigate and identify obstacles and avoid collision. The system identifies objects on left, right, front and down directions in real time. The system was designed using the ultrasonic sensor, IR sensor and Arduino Mega 2560. Using the three components, the system can be attached to the stick of blind persons and capable of sensing distance with the sensors and sends the distance data to the Arduino Mega2560 controller which then alerts the blind person using a buzz. They used Arduino IDE and C programming language for the software implementation of the system. It has the advantage of being low cost and implementable for developing countries. However, there is need to add to the system the ability to detect drop off on the way, ensure sufficient information of the obstacle ahead and safety of the visually impaired person.

According to Chandankhede and Kumar (2019) the concepts of Machine Learning and Artificial Intelligence has offered great algorithmic advantages to the development of real-time applications. Deep learning is inspired by the human neural system. Object recognition and computer vision are some of the applications of machine learning. Object and image detection are better handled with the help of deep learning techniques of AI which can be used to serve the blind impaired persons. Computer vision trains the computer to understand and perceive the visual world. They uncovered that CNN is the leading object detection that shows high level of accuracy. They proposed a deep learning technique to help a visually impaired person by capturing images in real-time, preprocessing, boundary detection, 2000 images restriction based on where the user visits regularly, image captioning then voice output. CNN and ReLU can be used for the system implementation using camera, speaker and any other hardware that could be mobile. The existing system are bulky, white cane with camera module were developed, but the system needs constant sweep rate with ground for proper functioning, cannot work in a crowded area, some existing system depends completely on sensors. The proposed system relies heavily on new methods for recognizing the training of the classes (that are images derived as a result of survey).

Durgadevi et al. (2020) proposed a machine learning system for blind assistance which was implemented using a camera, Raspberry pi and a speaker as hardware components. The system enables the user to detect an object using the camera based on the trained dataset and voice the name of the object using the speaker after sending it

to the Raspberry pi. They used a Neural Network model for the system. The system is cost effective simple and user friendly.

Legess and Seid (2020) proposed a CNN based model using a pre-trained model for the visually impaired persons to recognize Ethiopian currency banknotes in real time situations. 8500 images of Ethiopian currencies were collected as dataset. They evaluated the models with 500 real time videos under different views. They adopted Tensorflow object detection API, the faster-RNN and SSD MobileNet models. Transfer learning was used in the training of the datasets. Python language was used and opencv package was used for vision. For R-CNN 91.8% and for SSD MobileNet model has 79.4% accuracy. However, cannot detect fake currency notes.

Pathak and Aurelia (2019) proposed a mobile based smart currency recognition. The system was implemented in such a way that enables user to place their device against a currency and then the value of the currency can be known by giving the output as an audio signal and vibration formats using K-nearest neighbor and canny edge detection algorithm. The K-Means clustering algorithm to track points of interest on the currency, canny algorithm can be used to check for its edges of an object and finally use feature detection for valid currencies and perform currency matching by using feature matching algorithm available in MATLAB so as to finally show if the result if the image match accuracy is more than 75% else it requests the user to retake the image. The system was implemented only using Matlab and was not implemented in mobile, which appears more like a recommendation.

Yadav et al. (2020) are with the opinion that despite the availability of credit cards, electronic banking, internet, payment platforms, but still cash transactions are still going on because of its convenience and simplicity especially in rural areas where such transactions are unavailable. People with visual impairments suffer from identifying various currency values. Currency recognition can be of greatest help to the visually impaired persons. They proposed a currency recognition system for Indian rupees based on FAST and rotated YOLO V3 algorithm. They used six kinds of paper Indian currencies. The proposed system takes input of a given image, preprocesses it and coverts it from RGB to grayscale then a sober algorithm is applied for extraction of inner and outer edges. YOLO V3 algorithm is then used for clustering where clusters of features are made one by one, then it can recognize the feature of the image as either 200, 500, or 2000 depending on the input by comparing the same with the pre-trained dataset with the help of the YOLO V3 algorithm. The template matching is done using SURF key point detector on windows. The proposed system is camera based trained using image processing techniques on Indian currency dataset. The work can be extended to apply the classification to compare the original or counterfeit currency. It is possible to add foreign languages that can be used worldwide and to develop a recognition of currency notes on a low-end mobile phone for Visually Impaired persons and notify the user by voice note in regional language. It can be extended to recognize foreign currencies.

According to Mallikarjuna et al. (2021) many visually impaired persons in our society today face a lot of challenges and one of these challenges is object recognition. They can't navigate on their own, read and often feel isolated from their own community. There are exiting systems dressing such problems, but they have their own limitations. They proposed a solution to the problems faced by visually impaired by proposing a cost-effective system designed and implemented using IoT, machine learning and embedded technologies. The system was trained using Tensor Flow framework, camera to capture the image and a speaker to voice out the name of the object. A Raspberry pi was used to implement the system to which the camera sends the image after capturing, then it classifies the image then the type of the image is voiced out to the user via a speaker. Raspberry pi is trained using Tensor Flow machine learning framework for the classification of real-world object in python programming language. The proposed system helps identify four objects Car, Cat, Bus and Human being. They collected a set of 1615 images which forms the dataset. They used CNN algorithm and OpenCV software for pre-processing. The system is limited to object recognition only and the camera captures the image in one direction also, cannot work in the night. Therefore, the system can be extended for visual inspection of goods in industries, making a bionic eye.

The reviewed studies in this section demonstrate several approaches to aiding visually impaired individuals in currency recognition. Samant et al. (2020) and Almu and Muhammad (2017) showcase early systems using traditional image processing and manual feature extraction, which provide a low-cost solution but are limited by scalability and robustness. More recent works (e.g., Suranya et al. (2020) and Ng et al. (2020)) introduce deep learning often through transfer learning with CNN architectures to improve accuracy and processing speed. However, even these methods sometimes rely on limited datasets or face challenges when differentiating visually similar denominations. Overall, while deep learning methods have significantly increased recognition performance with many reporting accuracies above 95%, there remains a gap in addressing issues such as counterfeit detection and adapting to real-world variances like worn-out notes or varied imaging conditions.

2.4.2 Counterfeit Currency Detection

Bhatia et al. (2021) proposed a fake currency recognition using K Nearest Neighbour (K-NN) followed by image processing. They collected a high-quality image of both fake and genuine currencies as datasets using an industrial camera with 400x400 pixels dimension. They used a wavelet transform to extract features from the collected images. The attributes gathered after that includes Variance, skewness, kurtosis, Entropy, class of the currency. The data set has a total of 1372 images, out of which 610 are genuine and 762 are counterfeit currencies respectively. The further normalized the data to avoid biased classification due to some features on the currencies using MinMax Scaler which was imported from the sklearn. They trained the models using three (3) different algorithms including K-Nearest Neighbour (KNN) with the accuracy of 99.7%, Support Vector Classifier (SVC) with an accuracy of 97.5% and Gradient Boost Classifier (GBC) with 99.4% accuracy respectively. After the comparison, the KNN which has the highest accuracy was considered best algorithm for detecting counterfeit currencies based on the small amount of dataset used which may not be suitable for large datasets due its mode of operation which can be time consuming. As a future work, they proposed a Deep learning algorithm like CNN which can be suitable for large datasets and increasing the number of datasets. More so, more datasets should be added with high quality of real-world images and the use of convolutional neural network which have high accuracy in image processing scenarios. With the CNN the concept of wavelet transform can be removed as the CNN can analyse images from the input. Using CNN can also make the system more convenient and user friendly to use. In this case, the system can be improved and deployed for use by visually impaired persons by adding a voice note to detect the value and to distinguish between fake and genuine currencies.

According to Ogbuju et al. (2020) technological advancements has brought about the increase in cases associated with counterfeit currencies around the world. In Nigeria, the issue of counterfeit currencies is the biggest challenge faced in cash transactions. Therefore, it becomes important to utilize automatic means of counterfeit currency detections using machines to detect these fake currencies. They proposed a system that would help detect counterfeit banknotes and would contribute in curbing the menace of currency counterfeiting. To achieve this, they applied a deep learning approach using Faster Region Recurrent Neural Network (FRCNN) a class of CNN with 24 NN and 2 fully connected layers making a total of 76 nodes all together on the network to develop a naira detection model in Google Colab which was deployed to a mobile application called Real Naira. The system was tested with four (4) higher denominations of ₦100, ₦200, ₦500 and ₦1000 currency notes. They used a total of 730 images of the four (4) higher denominational currency notes. They extracted some relevant features that can be found on genuine currency banknotes and used these features for classification to identify the fake and genuine banknotes. The system is limited to detecting only the four higher currency banknotes using small number of datasets. Therefore, more datasets need to be added, with the addition of the four (4) lower currency banknotes ₦5, ₦10, ₦20, and ₦50 into system and possibly expanding the scope of the system to detect other African currencies.

According to Sarfaz et al. (2019) the advancement in digital imaging technology which includes the ability to print highly coloured papers has made it possible to produce fake currency Banknotes. The fake banknotes caused loss to anyone especially those involved in cash financial transactions. Therefore, there is a need to verify these currencies for smooth financial transactions. They proposed a cognitive computation-based approach for paper currency verification. They performed a scanning of both fake and original currencies to determine what both are made of using Scanning Electron Microscopy (SEM) and X-Ray Diffraction (XRD) analyses of counterfeit and genuine banknotes were performed. They used Support Vector Machine for classification based on the features like printing ink, chemical composition, and surface coarseness which they found a significant difference between the fake and the original currency. For experimentation and evaluation of performance purposes, they collected 195 Pakistani banknote images with 35 counterfeit banknotes. The system achieved 100% verification accuracy for well captured images. 60 genuine and 10 counterfeits of 500, 1000 and 5000 PKR each with threshold value of 0.4 were used for performance assessment and the system achieved 100% accuracy. However, more datasets should be added and can be deployed into web or mobile and another algorithm can also be used.

Rajebhosale et al. (2017) proposed a system based on image processing technique to identify fake currency notes automatically. They collected a dataset of 100, 500 and 1000 rupees using a camera, and they performed feature segmentation and template matching. Image processing was used to extract small parts of the images which matches the template image using template matching. Then they finally they deployed the template matching algorithm into a web application which enables the detection of both genuine and fake Indian rupees for the three denominations used. The goal of their work was to develop a friendly web application that can detect and recognize fake and genuine Indian rupees. They further propose that the method can be adopted and be used for real time recognition. The web application provides an interface to upload a currency note that the user would

want to verify, and the system displays the uploaded image(input) alongside the processed(matched) version of the image with the statues of the image either being original or fake under the image. The technique can be very adaptive and can be implemented in a real time world.

Agasti et al. (2017) proposed an Indian currency recognition using image processing. The proposed system gave an approach to determine and verify fake Indian currency notes by extracting various features of the currency using MATLAB. The new system has advantage of simplicity and high speed. They used the concept of edge detection to identify and detect certain features and image segmentation which divides the images into sub-regions that could be used for feature matching and distinguishing the features of fake and genuine banknotes. They itemized the features which can be used to identify currencies as follows: security threads, serial number, latent image, watermark, identification mark, etc. The process involves image acquisition, grey scale conversion and edge detection, image segmentation, feature extraction and calculation of its intensity the finally, if the condition is satisfied with 70% threshold, then it is considered as genuine otherwise classified as fake.

Zhang (2018) proposed a Single Shot MultiBox Detector (SSD) model which was based on deep learning as a framework and employing the convolutional neural network (CNN) for feature extraction of paper currencies. They used two models for comparing the MobileNet and Faster R-CNN. However, they found out that the CNN performed suitably well for currency identification requirements. Even when a currency is tilted, moved front or back, it can still be identified. By using CNN and SSD with accuracy of 96.6%. They collected a total of 300 raw images of New Zealand dollars with 50 images of 6 denominations each. However, the data was not sufficient, so they performed data augmentation by rotating, resizing, randomly zooming the images and the total number of images increased to 300x25 which equals 7500 images. They created a 6-layer CNN model and chose to use quadrilateral box and set the initial weights to 1.0 for the currency recognition training. They finally compared and analyzed the three models they used which includes MobileNet, Faster R-CNN and SSD. The SSD model was the most accurate for currency recognition. In the future, there is a need to add many different country currencies, use serial number of the currencies and surface patterns, use other deep learning models such as RestNet-101, Inception_V2 model etc.

In this section, the focus shifts toward distinguishing genuine from counterfeit notes. Bhatia et al. (2021) used a K-Nearest Neighbour (K-NN) approach with wavelet-based feature extraction to achieve very high accuracy, yet this method may be computationally intensive and less scalable to larger datasets. Ogbuju et al. (2020) leverage deep learning using Faster R-CNN to detect higher denominations in Nigerian currency, but their approach is constrained by a small dataset and is limited to only a subset of denominations. Sarfaz et al. (2019) employed advanced imaging techniques (SEM and XRD) with SVM classification, achieving perfect accuracy in controlled conditions, though their reliance on high-quality, specialized imaging limits practical deployment. These studies indicate that while high accuracy is achievable in counterfeit detection, real-world application is hindered by dataset quality, processing demands, and the challenge of integrating such systems into user-friendly platforms.

2.4.3 Currency Recognition and Detection Using Deep Learning

Bhutada et al. (2020) proposed a currency detection system for staff working at the forex bank to identify currencies of various countries around the world. They used a machine learning technique such as image processing to help detect the origin, name, and denomination of various currencies around the world. As there are many currencies, they developed a system that can take a particular currency as an input, pre-processes the image to extract the Region of Interest (ROI) from the notes which includes the origin of the currency and denomination based on some features like dimension, pigment and text clipping. They deployed the model into web using Wampserver, Django and python. The results obtained show that the model was able to detect currencies with 93.3% accuracy. The system identifies the currency using template matching of the original currency. The value and denomination of the currency can be identified based on the color, size and text colors used to separate the various currencies. However, few country currencies tested. In the future there is a need to progress to maximum currencies.

Sindhu and Varma (2020) proposed a currency recognition by using image processing. They developed an automatic currency recognition system using digital image processing method. It helps users to recognize details on a particular currency such as currency value and currency name by using the main characteristics on a currency. The targeted currency includes the Indian Rupees and the US dollars. The system uses image processing techniques to extract information from a currency image as an input and then match it with template images. NumPy and OpenCV in python are two frameworks used in this work to perform image processing functionalities and Thinker for applet development of the application. The system accepts input as an image, performs template matching and processing them displays an output. However, the trained dataset is low therefore more improvements need to be made and real time currency detection should be made where the user detects the

currency in real-time. There is a need to add more currencies as there are up to 180 currencies in 195 countries around the world.

Chakraborty et al. (2020) investigated various methods involved in currency recognition along with other techniques which can be applied in the process of identifying and recognizing currencies. They reviewed papers in order to identify recent developments in paper currency recognition. According to them, significant progress has been recorded over the years on currency recognition and detection as monetary transactions are inevitable part of our lives. Many people especially visually impaired persons who were most at times being deceived and being cheated because they cannot see. They pointed out some potential applications of currency recognition such as visually impaired assistance, counterfeit currency detection, Automatic selling goods, banking applications etc. though considerably much work has been done on currency recognition and detection, but there are still much vast opportunities to pursue in the future. Most of the models used for these works are Artificial Neural Networks (ANN) models with other kinds of neural networks such as Feed Forward, network, Back Propagation Neural Networks, Ensemble Neural Network. The framework of the existing methods is described, and the focus is on image acquisition, image localization, feature extraction, template matching and validating the output.

According to Linkon et al. (2020) an automatic detection and recognition of currencies can be very important for both the visually impaired persons and the bank because it can provide effective management for handling various paper currencies. They proposed an automatic approach for the detection and recognition of Bangladeshi Banknotes using a lightweight CNN architecture combined with transfer learning. They used ResNet152v2, MobileNet, and NASNetMobile as base models with two different datasets of Bangladeshi banknote images with one having 8000 images and the other 1970 images. They measured the performances of the models using the two datasets and obtained a maximum accuracy of 98.88% on 8000 image dataset using MobileNet, 100% on the 1970 images dataset using NASNetMobile, and 97.77% on the combined dataset (9970 images) using MobileNet. As a future work, increase the dataset, adjust the background of the images. Integration of fake currency banknote detection algorithms with the lightweight models.

Previous studies revealed that object recognition and detection using machine learning and deep learning has recorded a lot of progress. However, a model to recognize currency naira notes for four (4) higher denominations with the ability to distinguish between a genuine and a counterfeit note was proposed by Ogbuju et al. (2020). However, the work does not cover the recognition of lower currency naira notes. This work seeks to cover the currency recognition for eight (8) different naira currency denominations and voice out the values in English audio form using a mobile application.

This section reviews works that primarily apply deep learning techniques for recognizing and detecting currencies. Bhutada et al. (2020) and Sindhu and Varma (2020) explored models that combine image processing with machine learning to recognize multi-national currencies. Chakraborty et al. (2020) offer a broad review of deep learning methods, emphasizing that CNN-based models often enhanced through transfer learning provide the highest recognition rates. Linkon et al. (2020) further illustrates that lightweight CNN architectures can achieve high accuracy up to or above 98% and are suitable for mobile applications. Nonetheless, a recurring limitation is the need for large, diverse datasets to ensure model robustness and generalizability, especially under variable real-world conditions such as different lighting, angles, or wear of banknotes.

Across sections 2.4.1, 2.4.2 and 2.4.3, the literature reveals a clear trend: deep learning techniques, particularly CNNs augmented by transfer learning, have advanced the field of currency recognition, significantly outperforming earlier, traditional image processing methods. However, critical challenges remain, which includes

i. **Dataset Diversity and Quality:** Many studies rely on controlled or limited datasets, which may not fully capture the variability seen in real-world scenarios.
ii. **Computational Efficiency and Scalability:** While deep learning models achieve high accuracy, they often require substantial computational resources, making deployment on low-power or mobile devices challenging.
iii. **Robustness and Real-World Application:** Variations in image quality, lighting, and currency wear can adversely affect performance. Moreover, many systems do not yet robustly handle counterfeit detection alongside genuine currency recognition.
iv. **Integration for Practical Use:** Although high accuracies are reported in academic settings, translating these systems into accessible, user-friendly applications, especially for visually impaired persons, requires further work in terms of integration, real-time performance, and multi-language support.

These insights underline the need for future research to focus on building robust, lightweight, and scalable models that not only achieve high accuracy under controlled conditions but are also adaptable to the unpredictable variables of real-world usage.

3 Methodology

The methodology used for this work is deep learning. Many deep learning algorithms exist which have been used in object detection and classification. The most used which have proven to be more effective in vision-based systems and image processing systems is the convolutional neural network (CNN). The CNN concept has been applied in many real-world systems such as object recognition and detection, currency identification and classification, plant disease classification etc. Therefore, for the purpose of this work, a Deep Learning model based on transfer learning for Naira currency detection and recognition using MobileNetV2 CNN architecture was used after which the model was deployed into a mobile application using android studio for use on android mobile device.

3.1 System Flow Diagram

The system framework shows the processes involved in the development of the system from the start to the end. The process involved is captured in Figure 3.

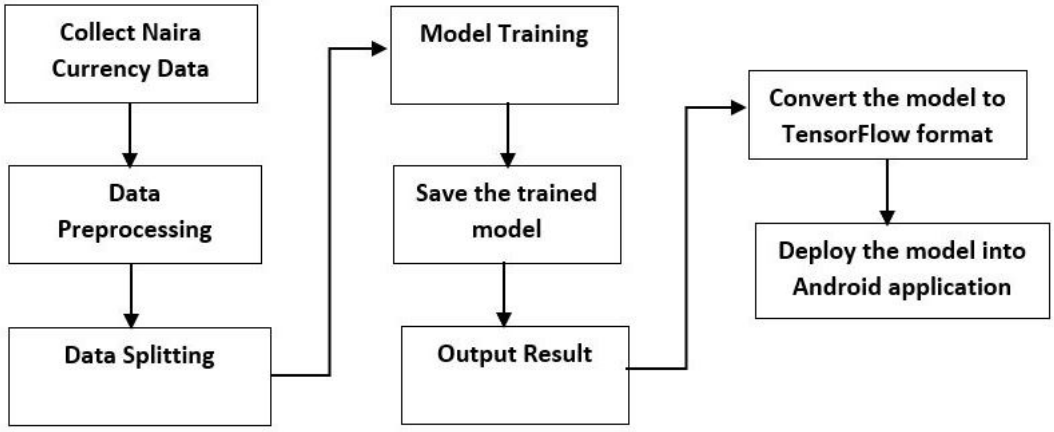


Figure 3: System Flow diagram

Figure 3 Shows the steps that is be followed to develop the system from getting the dataset, pre-processing, training, evaluation of training results and deployment into an android application for the user.

3.2 Dataset

The dataset was gathered using a mobile phone camera of 48Mega Pixels manually controlled, the dataset consists of eight denominations of Nigerian naira currency which includes: 261 five-naira denomination, 222 ten-naira denomination, 442 twenty-naira denomination, 269 fifty naira, 195-hundred-naira denomination, 362 two-hundred naira denomination, 572 five hundred naira denomination, 692 one-thousand naira denomination and 600 images of different objects for a non-currency class, making a total of 3615 images of naira notes based on the availability of the different currencies. The dataset collected are of 9 classes with each of the denominations making one class and the non-currency class, the images were named from 0, 1, 2, 3, 4, …, last. These images were resized to 224x224 pixels to make the suitable input dimension for the MobilenetV2 CNN architecture. The data was divided into 80% for training and 20% for testing respectively. The data augmentation method utilized includes the automatically generated augmentation while training the model and manual cropping of some of the data. The dataset used consists of nine classes, which were divided into 80% for training and validation and 20% for testing, making a total of 3615 images.

3.3 Training Environment Setup

The model was trained using python language in Spyder anaconda IDE. The trained model was converted to TensorFlow lite format making it ready to be deployed to a mobile device in the same Spyder anaconda IDE, the training was done on Windows 10 computer with 8GB RAM and 500GB HDD. The android app was designed and coded in android studio using Java programming language which was published to an android device for use.

3.4 System Model and Technologies

Figure 4 shows the model deployed into a mobile application which can detect a naira currency note and echo the value to the user. Python programming Language, MobileNetV2 CNN Architecture based on Transfer Learning, TensorFlow deep learning framework for training and Android studio for deploying the model into a mobile application were used.

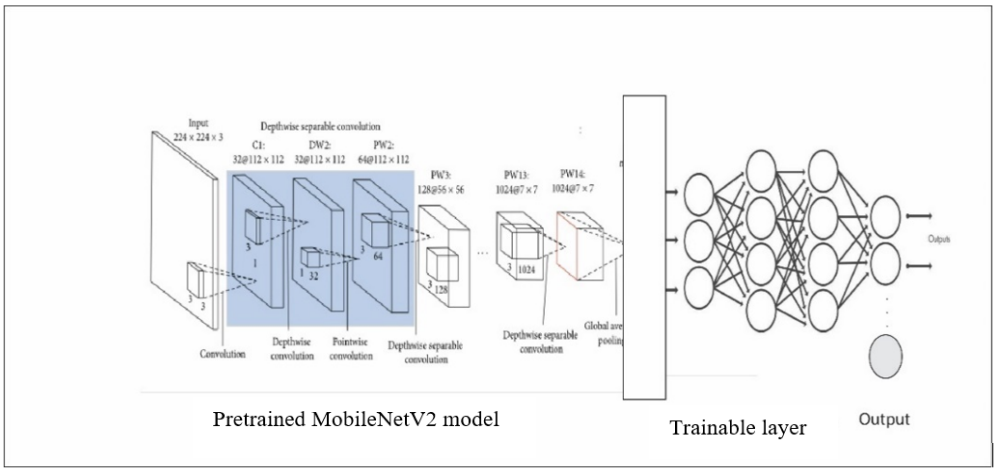


Figure 4: System Model overview (Howard et al., 2017)

Figure 4 shows how the MobileNetV2 pretrained CNN architecture used as a base upon which a layer containing the collected dataset is attached and trained using the concept of transfer learning. The first part is the fixed pretrained MobileNetV2 CNN architecture that is trained on ImageNet while the second part is the trainable layer which consists of the naira currency dataset with the input size of 224x224x3 attached at the output layer of the MobileNetV2 CNN architecture.

4 Results and Discussion

4.1 Results

The following results obtained from the training of the model using MobileNetV2 CNN architecture with 60 epochs (iterations) on the collected and preprocessed dataset. The pictural representation of the classification accuracy and cross entropy loss is shown in Figure 5 and Figure 6 respectively.

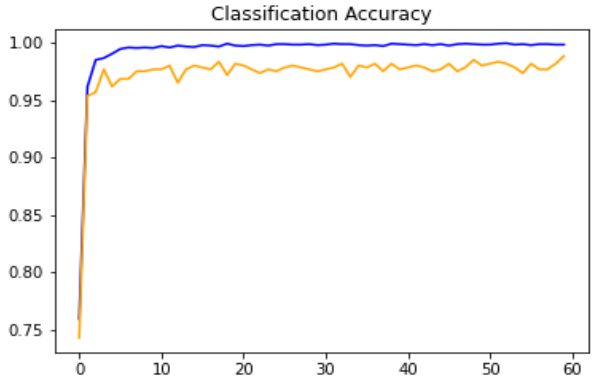


Figure 5: Classification Accuracy

Figure 5 shows the classification accuracy of the trained model. This accuracy is achieved after sixty (60) epochs.

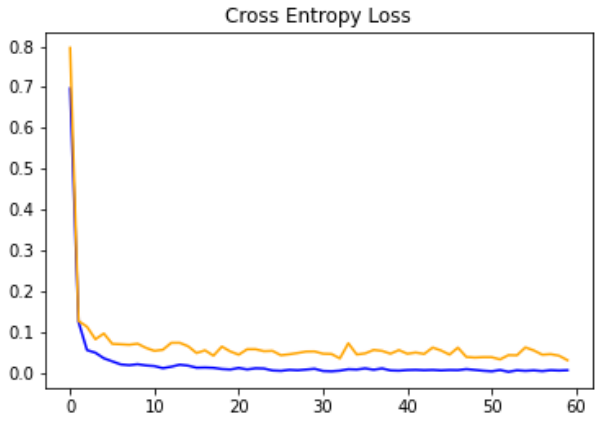


Figure 6: Cross Entropy loss

Figure 6 shows the cross-entropy loss of the trained model. This is the loss due to data imbalance and some factors associated with the trained model.

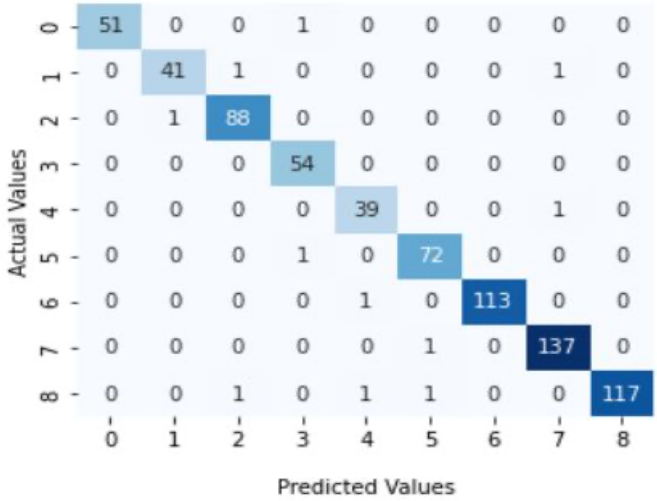


Figure 7: Confusion matrix for currency model

The confusion matrix in Figure 7 provides insight into the model's performance across different currency denominations based on the test data.
In the confusion matrix, each row represents an instance in a predicted class, and each column represents an instance in an actual class (Beheshti, 2022).

Table 2 shows the meaning of the list of classes in the confusion matrix in Figure 7.

Table 2: Classes and Meaning from Figure 7

Classes	**Interpretation**
0	Five (5) naira notes
1	Ten (10) naira notes
2	Twenty (20) naira notes
3	Fifty (50) naira notes
4	Hundred (100) naira notes
5	Two Hundred (200) naira notes
6	Five Hundred (500) naira notes

7	One Thousand (1000) naira notes
8	Noncurrency

In Figure 7, which presents the confusion matrix, the confusion matrix is used to ascertain the performance of the model as depicted in Table 3.

Table 3: Summary of the Performance Metrics

Class	Precision	Recall	F_1 – Score	Macro F_1 – Score	Accuracy
0	1.00	0.98	0.99		
1	0.98	0.95	0.96		
2	0.98	0.99	0.98		
3	1.00	1.00	1.00		
4	0.95	0.97	0.96		
5	0.97	0.99	0.98		
6	1.00	0.99	1.00		
7	0.99	0.99	0.99		
8	1.00	0.97	0.99		
				0.98	98%

The model achieves high precision and recall across all currency denominations, suggesting strong learning of distinguishing features. The model performed well across all classes, with precision and recall values ranging between 0.95 and 1.00, confirming its robustness. Focusing on class specific metrics, Class 3 which is Fifty Naira (₦50) achieved perfect classification, indicating that its unique features are well captured by the model.

Lower accuracy in non-currency classification suggests some non-currency objects may have textures or colors similar to currency notes which resulted in minimal missclassifications. The model performs exceptionally well, with F1-Scores above 0.95 for all classes. The results suggests high reliability in real-world applications. Though there are missclassifications in some of the classes, however, this is open to improvement.

4.2 Deployment Results

The model was deployed into a mobile application from Android studio and the results shown in figures 8 and 9. The model was trained and saved as model.h, and converted to Tensorflow lite format currency model.tflite using python language in Spyder IDE, the mobile application was coded using Java programming language.

This section shows results obtained from the model deployment and how it works on a mobile device. The system automatically opens the camera immediately it is launched, after which the user holding a naira currency can capture the image using the android's volume down button, then the system detects and echoes the naira currency value in audio form in English Language.

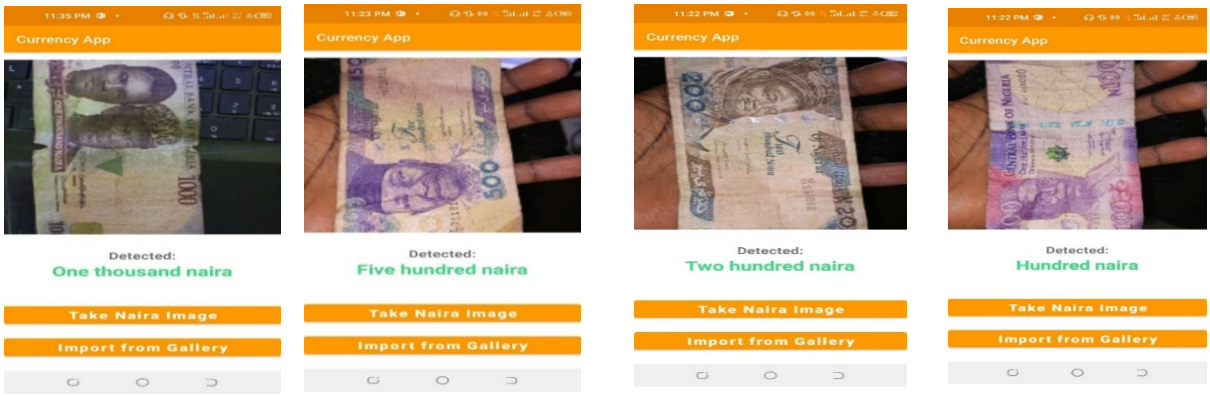


Figure 8: Detected images of one thousand, five hundred, two hundred and one hundred naira notes only

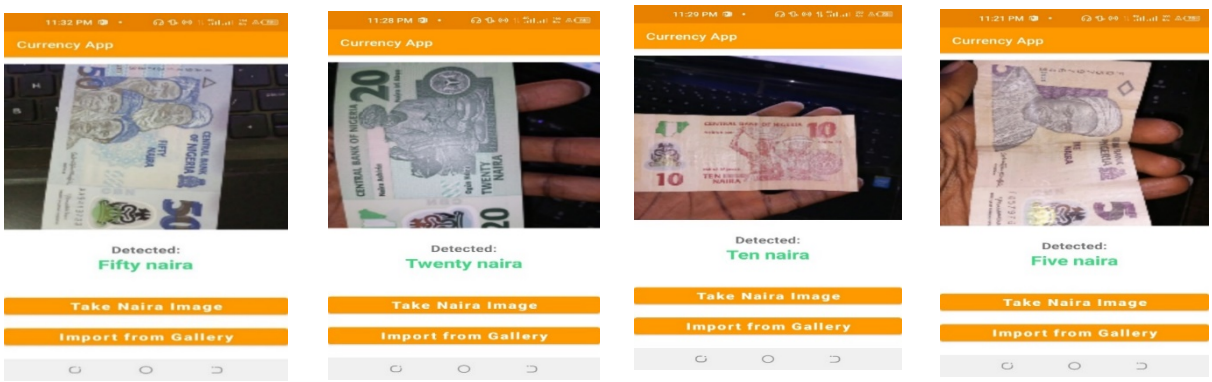


Figure 9: Detected images of fifty, twenty, ten and five naira notes

The results obtained from the training of the model using the mobileNetV2 CNN architecture is 98% accuracy. This indicates that the model's training accuracy is good. Also, the performance metrics shows that the model's performance is good. Most classes achieve recall (accuracy) levels between 97% and 99%, with Class 3 (Fifty naira notes) reaching 100%. The slight drop for Class 1 which has 95% accuracy suggests that Ten naira notes might be more challenging to distinguish, possibly due to visual similarities with adjacent denominations or variations in the note's appearance. Similarly, Classes 4 and 8 both at 97% accuracy shows a small degree of misclassification that might be due to subtle variations in design or imaging conditions.

Overall, the confusion matrix reveals an excellent classification performance across all classes, with only minor misclassifications that are within acceptable limits for real-world applications. This high level of accuracy indicates that the model is very reliable in recognizing the different currency notes and distinguishing them from noncurrency images.

Achieving such high accuracy suggests that the model have been trained on a well-balanced dataset, which suggests that its generalization to real-world conditions can be effective, though some of the classes have fewer datasets, however, automatic data augmentation has been used to increase the number of datasets during training. MobileNetV2 CNN Architecture have shown effectiveness in feature extraction even with few number of datasets since it is a pretrained model. Moreover, the model does not account for worn-out currencies which may affect real-world usability in such cases.

The deployment into the mobile application as seen in the results worked as expected and can be installed on mobile devices for detection of naira notes and voicing out the value of the corresponding classified naira currency in audio form in English Language. This work has added to the existing system more features such as dataset and audio integration which makes it suitable to be used by visually impaired persons. The system can also be used to teach about naira currency for children.

5 Conclusion

Looking at the state of art today, visually impaired persons need assistance to carry out their day-to-day activities especially when it has to do with financial transactions. This research explored the given problem and tries to address this problem which the visually impaired persons are faced with. Therefore, this research focused on the development of a model for the recognition of eight (8) denominations of Nigerian naira currency using deep learning CNN architecture MobileNetV2 model based on transfer learning for visually impaired persons. A total of 3615 images datasets were collected for the training of the model, the collected dataset was pre-processed, divided into 80% training and 20% testing, and the model was trained, saved, and converted into the TensorFlow format for deployment. The trained model achieved an accuracy of 98% which was trained using python programming language in the Spyder Anaconda IDE, the model was deployed into a mobile application which can be used by the visually impaired persons for the recognition of naira notes. The mobile application captures the currency image using the mobile phone's camera, then the captured image is sent to the currency model running at the application's background and the model classifies the currency image to its denomination. The mobile application then outputs the classified currency value both in text and in audio form. The user can listen to the value of the currency echoed by the mobile phone's speaker.

The system classifies only eight Nigerian naira currency denominations (5-1000). The work can further be extended to include the classification of Nigerian coins. Moreover, the result of currency classification voices out the values of Nigerian naira currency in English language, the work can be extended to add multiple languages used in Nigeria so that it can be used to learn about the naira currency in major Nigerian languages and for wider benefit and coverage.

References

Agasti, T., Burand, G., Wade, P., & Chitra, P. (2017, November). Fake currency detection using image processing. In IOP conference Series: materials Science and Engineering VIT University, Vellore, Tamil Nadu, India, 263(5), 1-8. IOP Publishing.

Akano, O. F. (2017). Vision health disparities and visual impairment in Nigeria: A review of the Nigerian National Blindness and Visual Impairment Survey. African Vision and Eye Health (AVEH): 76(1), 1-5. DOI:https://doi.org/10.4102/aveh.v76i1.345

Almu, A., & Muhammad, A. B. (2017). Image-Based Processing of Naira Currency Recognition. Annals, Computer Science Series: 15(1), 169-173.

Beheshti, N. (2022). Guide to Confusion Matrices and Classification Performance Metrics. Retrieved on 21st July, 2022 from https://towardsdatascience.com/guide-to-confusion-matrices-classification- performance metrics-a0ebfc08408e

Bhatia, A., Kedia, V., Shroff, A., Kumar, M., Shah, B. K., & Aryan. (2021). Fake Currency Detection with Machine Learning Algorithm and Image Processing. IEEE Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems Vaigai College Engineering (VCE), Madurai, India: 755-760. DOI: 10.1109/ICICCS51141.2021.9432274.

Bhutada, S., Reddy, V. K., Chevva, A., Kadapala, S. & Gella, M. (2020). Currency Recognition Using Image Processing. *International Journal of Scientific Development and Research (IJSDR),* 5(4), 275-278.

Chakraborty, K., Basumatary, J., Dasgupta, D., Chandra J. K., Mukherjee, S. (2020). Recent Developments in Paper Currency Recognition System. *International Journal of Research in Engineering and Technology (IJRET)*, 2(11). Available @ http://www.ijret.org222 ISSN: 2319-1163.

Chandankhede, P. & Kumar, A. (2019). Deep Learning Technique for Serving Visually Impaired Person. IEEE 9th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-19) Nagpur, India: doi: 978-1-7281-3506-9/19, 1-6.

Chaubey, N., Parikh, S., & Amin, K. (Eds.). (2020). Computing Science, Communication, and Security. Communications in Computer and Information Science. doi:10.1007/978-981- 15-66486

Daniel, U. (2020). Nigeria Association of the Blind (NAB). Retrieved on 25th October, 2021 from https://www.nigeriaassociationoftheblind.org/

Durgadevi, S., Thirupurasundari, K., Komathi, C. & Mithun, S. B. (2020). Smart Machine Learning System for Blind Assistance. IEEE International Conference on Power, Energy, Control and Transmission Systems (ICPECTS): 1-4 DOI: 10.1109/ICPECTS49113.2020.9337031

Gamage, C. Y., Bogahawatte, J. R. M., Prasadika, U. K. T., & Sumathipala, S. (2020). DNN Based Currency Recognition System for Visually Impaired in Sinhala. IEEE 2nd International Conference on Advancements in Computing (ICAC) Colombo, Sir Lanka: 1, 422-427. DOI: 10.1109/ICAC51239.2020.9357295

Geekforgeeks. (2019). Unified Modeling Language (UML) | An Introduction. Retrieved from https://www.geeksforgeeks.org/unified-modeling-language-uml-introduction/ on 21st July, 2022.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wand, W., Weyand, T., Andreetto, M. & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: 1704.04861

Islam, M. T., Ahmad, M., & Bappy, A. S. (2021). Real-Time Bangladeshi Currency Recognition Using Faster R-CNN Approach for Visually Impaired People. In Communication and Intelligent Systems:147-156. Springer, Singapore.

Jafri, R., Abid, S. A., Arabnia, H. R. & Fatima S. (2013). Computer Vision-based Object Recognition for the Visually Impaired in an Indoors Environment: A Survey. Vis Comput Springer: DOI: 10.1007/s00371-013 0886-11 26

Legess, M. M. & Seid, H. W. (2020). Real-Time Ethiopian Currency Recognition for Visually Disable Peoples Using Convolutional Neural Network. Research square: 1-24. DOI: https://doi.org/10.21203/rs.3.rs125061/v1.

Linkon, A. H. M., Labib, M. M., Bappy, F. H., Sarker, S., Jannat, M. E., & Islam, M. S. (2020). Deep Learning Approach Combining Lightweight CNN Architecture with Transfer Learning: An Automatic Approach for the Detection and Recognition of Bangladeshi Banknotes. 2020 IEEE 11th International Conference on Electrical and Computer Engineering (ICECE) Buet, Dhaka, Bangladesh: 214-217.

Mallikarjuna, G. C. P., Hajare, R., & Pavan, P. S. S. (2021). Cognitive IoT System for visually impaired: Machine Learning Approach. Materials Today Proceedings. 1-7. doi: 10.1016/j.matpr.2021.03.666

Mathworks (2021). Convolutional Neural Network. Retrieved on 11th November, 2021 from https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html

Naing, M., Nwe, N. O., & Thwe, M. O. (2019). Smart Blind Walking Stick using Arduino. International Journal of Trend in Scientific Research and Development (IJTSRD):3(5) August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470, 1156-1159.

Ng, S. C., Kwok, C. P., Chung, S. H., Leung, Y. Y., & Pang, H. S. (2020). An Intelligent BanknoteRecognition System by using Machine Learning with Assistive Technology for Visually Impaired People. 10th IEEE International Conference on Information Science and Technology (ICIST) Bath, London, and Plymouth, United Kingdom: 185-193. doi:10.1109/icist49303.2020.9202087

Ogbuju, E., Usman, W. O., Obilikwu, P. & Yemi-Peters, V. (2020). Deep Learning for Genuine Banknotes FUOYE. *Journal of Pure and Applied Sciences (FJPAS)*, 5(1), 56-67. ISSN: 2616-1419.

Pascolini, D., & Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. British *Journal of Ophthalmology*, 96(5), 614-618.

Pathak, A., & Aurelia, S. (2019). Mobile Based Smart Currency Detection System for Visually Impaired. *Journal of Image Processing and Artificial Intelligence*, 5(3), 1–4. http://doi.org/10.5281/zenodo.3354669.

Rajebhosale, S., Gujarathi, D., Nikam, S., Gogte, P., & Bahiram, N. (2017). Currency Recognition System Using Image Processing. *International Research Journal of Engineering and Technology (IRJET)*, 4(3), 2559-2561.

Samant, S., Sonawane, S., Thorat, R., Shah, P. B., & Kulkarni, N. P. (2020). Currency Recognition System for Visually Impaired People. *International Journal of Advance Scientific Research and Engineering Trends*, 5(3), 31-35.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted Residuals and Linear Bottlenecks. Proceedings Of the IEEE Conference On Computer Vision And Pattern Recognition (Iccvpr) Salt Lake City, Ut, Usa, 4510-4520: 10.1109/Cvpr.2018.00474.

Saranya, K. S., Badhan, A. K., Alekhya, A., Madhumitha, C., & Charmika, V. D. (2020). Currency Counting for Visually Impaired Through Voice using Image Processing. International *Journal of Engineering Research & Technology (IJERT)*, 9(5), 195-199.

Sarfraz, M., Sargano, A.B., & UI Haq, N. (2019). An Intelligent System for Paper Currency Verification using Support Vector Machines. Scientia Iranica, 26(1) (Special Issue on: Socio-Cognitive Engineering), 59-71: doi: 10.24200/sci.2018.21194

Simske, S. (2019). Meta-analytics: consensus approaches and system patterns for data analysis. Morgan Kaufmann.

Sindhu, R. & Varma, P. (2020). Currency Recognition System Using Image Processing. *International Journal of Scientific & Engineering Research,* 11(4), 1595-1601.

Tasnim, R., Pritha, S. T., Das, A., & Dey, A. (2021). Bangladeshi Banknote Recognition in Real-time using Convolutional Neural Network for Visually Impaired People. 2021 IEEE 2nd international Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) American International University, Dhaka, Bangladesh: 388-393.

Veeramsetty, V., Singal, G., & Bidal, T. (2020). Coinnet: platform independent application to recognize Indian currency notes using deep learning techniques. Multimedia tools and applications, 79, 22569-22594

WHO. (14 October, 2021). Blindness and vision impairment. Retrieved on 25th October, 2021 from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

Yadav, S., Ali, Z., & Gautam, K. S. (2020). Currency Detection for Visually Impaired. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 7(5), 999-1002. www.jetir.org (ISSN-2349-5162)

Zhang, Q. (2018). Currency Recognition Using Deep Learning. A thesis submitted to the Auckland University of Technology in partial fulfilment of the requirements for degree of Master of Computer and Information Sciences (MCIS): I-95.

Zhu, W., Braun, B., Chiang, L. H., & Romagnoli, J. A. (2021). Investigation of transfer learning for image classification and impact on training sample size. *Chemometrics and Intelligent Laboratory Systems: An International Journal Sponsored by the Chemometrics Society*, *211*(104269), 104269. doi:10.1016/j.chemolab.2021.104269