VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2025, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 13, Number 1, January-March 2025 pp: 141-152

An Efficient System for Urdu Sign Language Recognition using Support Vector Machine(SVM), Convolutional Neural Network (CNN) and Ensemble Machine Learning (EML)

Kaneez Zainab Abbasi[1], **Abdul Khalique Maitlo**[1*], **Sila Zubair**[1], **Kainat Khero**[1], **Shazmina Khan**[1]

[1]Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST) University, Larkana, Pakistan

Keywords:*Support Vector Machine (SVM), Convolutional Neural Network (CNN), Ensemble model, Urdu Sign Language; Hand Gesture, Sign Language Recognition (SLR).*

Journal Info:
Submitted:
June 06, 2024
Accepted:
March 17, 2025
Published:
March 31, 2025

Abstract Sign language has significant problems in the everyday life of deaf and hard-of-hearing people. We have used a Support Vector Machine (SVM), a Convolutional Neural Network (CNN), and an Ensemble Machine Learning (EML) model that combines their outputs as our machine-learning technique. We seek to design a USL recognition system that will address communication gaps. Firstly, we reviewed the "Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet, through Machine Learning". The dataset has various characteristics, such as image quality, size, and class distribution. The dataset plays a pivotal role in training and evaluating the proposed models. It includes a diverse range of images representing the Urdu Sign Language (USL) alphabet, ensuring the models are exposed to varying hand configurations, backgrounds, and lighting conditions. This diversity helps improve the generalizability of the trained system. During preprocessing, we performed normalization, resizing, and augmentation techniques to enhance the robustness of the data and prevent overfitting. Results indicated that the ensemble approach outperformed the individual models, achieving higher classification rates for several challenging hand configurations. The developed system shows promising potential for real-world applications in bridging the communication gap faced by the deaf and hard-of-hearing community in Pakistan.

***Correspondence author email address:** a.khalique@lrk.szabist.edu.pk
DOI: 10.21015/vtse.v13i1.1818

1 Introduction

One of the most essential and fundamental requirements for social survival is communication. Deaf and hard-of-hearing people mostly use sign language as they communicate with one another. However, it is difficult for non-deaf people to understand what they are conversing about [1]. Communication is the most effective means of interaction among people, society, and communities. Nonetheless, for the deaf-mute community, the bridge is usually incomplete as sign language is their primary method of communication. It is worth noting that deaf and mute individuals are fluent in a unique language, sign language, and their language can be a source of isolation because many people do not get it. With growing interconnectedness worldwide, inclusivity, accessible communication, and access to all are essential. The development of sign language recognition systems aims to reduce communication obstacles and help deaf and hard-of-hearing people, as well as those who want to interact with them, to communicate freely. Our research project is directed at developing an effective USL recognition system. This study has enormous social and practical applications in Pakistan, where Urdu is one of the main languages.

1.1 Problem statement

Our research focuses on a critical matter: the formation of a reliable and effective system for USL recognition to consider the intricacy and diversity of USL. Urdu Sign Language is complex and involves complex hands and fingers. It is a technological challenge and a social requirement because of its importance for the deaf and hard-of-hearing community in Pakistan. This problem statement will focus on developing a system capable of translating USL gestures into text or audio, thus enabling the deaf and hard-of-hearing people to communicate efficiently with the hearing world. However, this problem statement underscores the role of technology in fostering access, promoting inclusion and meaningful communication between these two communities.

Our research aims to overcome these obstacles and provide a workable option whose primary goal is to promote communication between deaf and hard-of-hearing individuals.

1.2 Objectives and Hypothesis

The primary objectives of our research can be summarized as:

Creating a highly effective USL recognition system.

Studying and applying machine learning technologies, including SVM and CNN, to USL recognition.We are exploring the possibility of using ensemble learning to bring together different models' strengths and improve overall recognition accuracy.

We hypothesize that Using machine learning and an extensive dataset can result in the development of an effective USL recognition system. Our working hypothesis is that SVM and CNN could be fine-tuned separately for high recognition accuracies. An ensemble model that fuses the strengths of SVM and CNN will outperform single models, resulting in a highly accurate recognition system.

2 Literature Review

In recent years, there has been rapid development in sign language recognition due to the growing need for clear communication between people who are hard of hearing and other members of society. However, this section surveys several research papers on sign language recognition systems and their significance.

Expanding on this, recent research by [1] has contributed to the field by developing a PSL dataset and a system that identifies hand configurations of the Urdu alphabet. The application and dataset are imperative in reducing the communication barrier between the deaf and non-deaf communities.

In another research, the authors comprehensively described modern hand gesture-and-sign language recognition approaches in their review paper. The methods are divided into different phases, including data collection, preprocessing, segmentation, extraction of attributes, and classification. Thus, this comprehensive review focuses on the issues encountered in gesture research and their importance in sign language recognition [2]. Highlights the importance of multidisciplinary cooperation in developing sign language recognition, generation, and translation systems. The study reveals the need for collaborative efforts to address the multifaceted challenges of sign language processing. This integrative approach brings the complexity of the field and the possibility of multidimensional solutions to light [3]. Examines the role of mobile applications in improving individual deaf learning experiences. The study uses mobile technology to facilitate learning basic concepts using Pakistan Sign Language (PSL). It shows that mobile applications can significantly enhance learning and understanding of topics, thereby addressing the educational challenges of people who are hard of hearing. In the domain of hand gesture recognition [4].

Introduces a robust method using 3D Convolutional Neural Networks (3DCNN). The authors use transfer learning to overcome the limitations associated with limited labelled data sets. The survey shows that signer-dependent and signer-independent attributes are widely accepted [5]. Examines multiple sign language recognition and shows that combining computer vision and Leap Motion data significantly enhances recognition. This delayed fusion method complements the single-view classification of the image and leap motion data, highlighting the power of integrated features [6]. Highlighting the importance of using computer vision sign language recognition systems to communicate the communication gap between hearing and deaf people, the study includes existing methods and their impact story. It emphasizes the role of vision-based methods in this area [7]. Research by [8] provides valuable insights into the development of sign language machine translation. Their comprehensive review reviews 147 high-quality research articles and publications on sign language interpretation, classifies the methodologies used for each category, and discusses theoretical foundations. This work emphasizes the importance of methodology diversity in addressing sign language interpretation challenges. Another significant development is presented in the paper [9], which presents a method for recognizing standardized Pakistan Sign Language (PSL) characters using bare hands. This approach combines object vision-based learning with multi-kernel learning (MKL) and support vector machines (SVM) to achieve incredible accuracy, addressing the need for cost-effective and easy-to-consume methods and the role of the solution.

This work provides a comparative analysis of sensors used for sign language recognition. The review discusses sensor-based bright fingerprints and emerging trends, emphasizing the need for advanced sign language recognition techniques [10]. Provides an in-depth learning approach to Arabic Sign Language translation, addressing challenges of accuracy, gesture complexity, and high noise levels. Their proposed algorithm, combining convolutional neural network (CNN) and recurrent neural network (RNN), exhibits significant improvement in accuracy and efficiency [11]. Please contribute to the field with their research on sign language recognition and recognition using deep learning. Paradigms such as LSTM and GRU can recognize Indian Sign Language (ISL) gestures from video images, significantly reducing communication barriers [12]. Implement Pakistani Sign Language (PSL) recognition using Bag-of-Words (BoW) and Support Vector Machines (SVM). This vision-based algorithm uses image data and SURF annotations to achieve high classification accuracy, adding valuable features to the field [13]. In [14], a systematic review using machine learning has extensively studied Urdu Sign Language Classification. This study summarises available datasets and machine learning methods for Urdu Sign Language recognition, highlighting existing research status and areas for improvement.

Support the field with effective pose-based graph convolution methods for sign language recognition. Their model achieves improved accuracy on different data sets and offers a promising solution for identifying signed words [15]. Provides a desktop application to recognize American Sign Language (ASL) gestures in real-time and convert them into words. Their research focuses on improving recognition accuracy, addressing the growing need for effective sign language communication[16]. Provides interactive and model-based deep learning for Urdu Sign Language, addressing the communication gap between ordinary and differently abled people. Their model significantly improves descriptive accuracy and promises to prevent communication fragmentation

[17].

This comprehensive literature review covers various sign language recognition, region-specific recognition algorithms, methodologies, machine learning, and deep learning. These contributions collectively advance the field and provide insights into future research directions.

3 Proposed Methodology

We have utilized a quantitative methodology as shown in Figure 1. To achieve this, we must engage in quantitative data analysis that relies on machine learning algorithms. Our research has established a structured framework using support vector machines and convolutional neural networks to train and evaluate effective models.

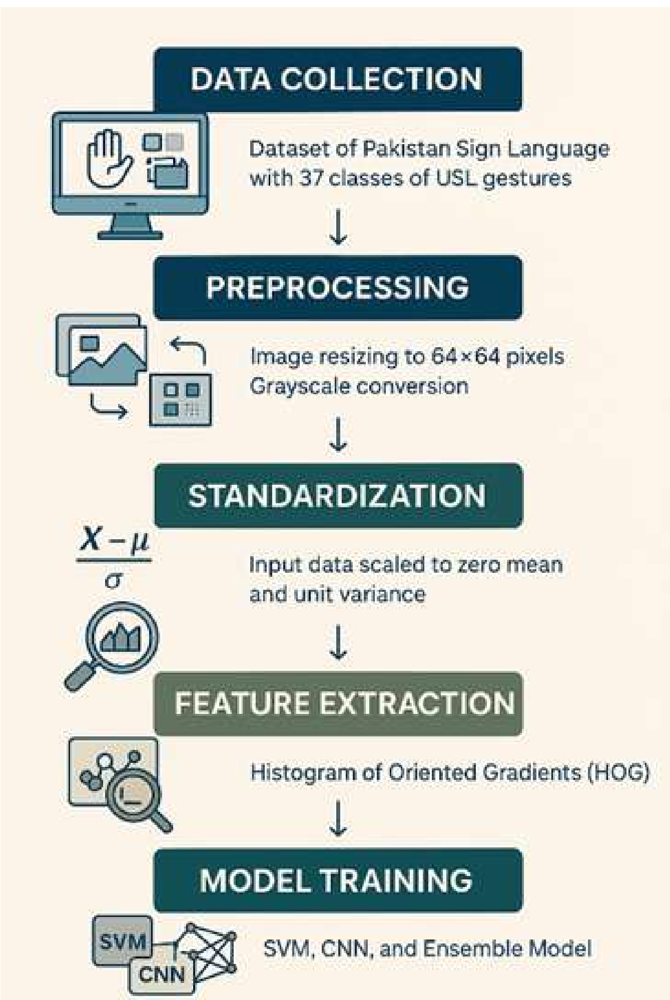


Figure 1. Proposed Methodology

3.0.1 Data Collection:

The primary data input for this study is the "Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet through Machine Learning" Dataset of Pakistan Sign Language [1]. This dataset includes a collection of USL gestures divided into 37 classes according to Urdu alphabets. It consists of a set of sign gestures, which makes it adequate for the training and testing of our models.

3.1 Preprocessing

In this study, two main preprocessing techniques are applied to the data set to increase the quality and accuracy of image input. These techniques are essential for machines' practical learning model training, especially for sign language recognition. A standard method for reducing the amount of data to be processed is to convert an image to a grey scale. All the sign gesture images undergo image preprocessing to conform to a single size and format and are converted into grayscale. CNN processing is preferred for grayscale images. Data standardization is applied to make sure that the input data are in the unified format that the model needs.

3.1.1 Image Resizing and Grayscale Conversion

The images in the dataset are converted to a fixed size of 64x64 pixels. This step ensures all models are identical and allows consistent input patterns for machine learning models. The choice of 64x64 pixels balances computational effort and preserves crucial visual information. This resolution was ideal for the Urdu Sign Language Recognition task, simplifying the subsequent task while preserving important information.

Grayscale adjustment is performed to convert a color image to a grayscale representation. Removing color information and representing images in different greys makes the data set more manageable and computationally efficient, which is paramount to recognizing sign language gestures. The decision to use grayscale instead of color images was mainly due to its ability to simplify the data while preserving characteristics important to our analysis.

3.1.2 Standardization

Data standardization is an essential preprocessing step, especially for machine learning models such as Support Vector Machines (SVM). This serves the purpose of ensuring that all feature values (pixel values in this case) display the same scale, thereby reducing problems associated with different numbers of features that can adversely affect model performance. The standardization process involves subtracting each object's mean (pixel value) and dividing by the standard deviation. This gives each feature zero mean and a standard deviation of one. Mathematically, this can be expressed as:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

where X represents the trait value, μ represents the mean, and σ the standard deviation. Standardization is essential in SVM because the algorithm works based on the distance between data points. Non-standardized data can lead to inaccurate distance estimates and, consequently, inaccurate sampling results. Standardization, therefore, ensures that the SVM converges faster, finds better solutions, and helps to increase model performance.

3.2 Features Extraction

The feature extraction method we used is the Histogram of Oriented Gradients (HOG). HOG is very good at tracing the specific patterns and forms in the sign gestures.

A framework of technical aspects was processed and discussed in various structures, including image preprocessing, principal component analysis (PCA) for dimension reduction, and support vector machine (SVM) for classification. This highlights the importance of histogram equalization to maximize recognition accuracy. An accuracy of 76.8% has been noted, and further research is suggested to improve the system's robustness and accuracy [18].

HOG takes gradient distributions in each patch of an image, encoding essential information about sign gesture contours. Another advantage of HOG is its invariance to illumination variations and minor image distortions. This robustness ensures that extractions remain reliable even in real-world situations where lighting or viewpoints may vary. As an integral part of our feature extraction process, HOG somehow deconstructs the image, building a mathematical representation that covers the essential characteristics of USL gestures. This advanced feature forms the basis for our machine learning models, increasing their ability to recognize the signal in more accurate gestures.

3.3 Model Training

Our research starts with the foundation phase, dedicated to training machine learning models to accurately recognize Urdu Sign Language (USL) gestures. In this critical section, we investigate the robustness of this training system, emphasizing the two main models, support vector machine (SVM) and convolutional neural network (CNN), which take center stage internally in our USL recognition system.

A deep convolutional neural network (CNN) algorithm for recognizing Arabic sign language recognition. Trained on a large data set of Arabic sign images, the system achieves an accuracy of 97.6%. The study demonstrates the potential of deep learning techniques in sign language recognition [19].

These models work together to provide comprehensive solutions for gesture classification and subsequent text generation. The journey begins with a support vector machine (SVM), a versatile tool known as classification robustness. It plays a vital role in determining the accuracy and efficiency of USL gestures. A continuously parallel convolutional neural network (CNN) begins its training journey. It is known for extracting spatial features directly from image data. These models, equipped with their unique features and strategies, ensure that the system can effectively interpret the distinctions of USL, facilitating its translation into human interaction.

4 Proposed Algorithm

4.1 Hyperparameter tuning for support vector machines (SVM) using GridSearchCV:

Urdu Sign Language (USL) gesture recognition requires the selection of an appropriate machine learning model and the careful optimization of its hyperparameters. In our research, we have used the GridSearchCV method as an important approach to fine-tune the hyperparameters of a Support Vector Machine (SVM). The proposed research outlines the development of a PSL recognition system. The system uses supervised learning in conjunction with SVM to classify static sign language gestures with an accuracy of 83% [20].

The main objective is to enhance the classification performance of the SVM, which ensures excellence in USL gesture recognition accuracy. This section provides an in-depth insight into the hyperparameter tuning process. GridSearchCV, a method embedded in hyperparameter tuning, provides a systematic and comprehensive method for searching the hyperparameter space of our SVM model the best of the best. The primary hyperparameters under examination are the selection of the kernel and the fine-tuning of the cost parameter 'C'.

4.1.1 Kernel selection: The core expression of the SVM

The SVM kernel contains a key that enables it to distinguish between different classes in a data set. Grid search requires two main kernel methods: Radial Basis Function (RBF) and linear kernel. The RBF kernel is well suited to capture complex decision boundaries in feature space, making it a good choice for complex USL gestures. On the other hand, the linear kernel is good at linearly separable data but could be more robust at decision-making. The grid search systematically searches both kernels, identifying the one that best matches the complexity of the USL signal.

4.1.2 Cost parameter 'C': Creates balance

The value parameter 'C' is essential in SVM classification accuracy. This is a regularization criterion, determining the trade-off between maximizing margins and minimizing classification errors between support vectors. In grid search, 'C' is carefully adjusted to achieve the best balance between these values. A low value of 'C' indicates a wide margin but can tolerate some misclassification. In contrast, a high value of 'C' indicates a more uniform distribution but may result in a narrower interval. The main objective of this hyperparameter tuning procedure is to increase the classification accuracy of the SVM model. By systematically searching various combinations of kernels and 'C' values, GridSearchCV proposes a hyperparameter scheme that results in the highest recognition accuracy for USL gestures. This careful optimization ensures that SVM performs the best task to distinguish between different sign gestures, thus improving the overall performance of the USL recognition system. In summary, using GridSearchCV for hyperparameter tuning in SVM models demonstrates our commitment to improving model performance. This advanced approach, coupled with kernel selection and 'C' parameter tuning, makes SVM compatible with the exceptionally complex USL gesture, highlighting its accuracy in classifying these complex hand gestures.

Our hyperparameter tuning procedure uses 18 different fits to a 3-fold cross-validation analysis phase. Processor requirements are different for each system. Notably, the structures with RBF kernels have longer loading times, with an average loading time of about 5.3 seconds, while the linear kernels require less computation and an average of 3.6 seconds. This difference arises due to the nature of the kernel function, where the RBF kernel introduces nonlinearity, increasing the computational complexity. After careful observation, the network search ends with the selection of the best SVM model. This optimized model significantly improves USL gesture recognition, reaching a validation accuracy score of 96.36%. This result highlights the effectiveness of hyperparameter tuning and its significant impact on model performance.

4.2 CNN Model Architecture

The model structure starts with the lowest input layer, which determines the input data size. Configuring the input layer correctly is crucial because it sets the phase for subsequent layers. After the input layer, the first explicitly defined layer is the dense layer. This layer contains 128 neurons using the Rectified Linear Unit (ReLU) processing function. ReLU introduces nonlinearity into the model, allowing the capture of complex patterns and relationships in the data.

Moreover, after this dense layer, a dropout layer is added at a rate of 0.5. Dropout is a regularization method that resists overfitting by randomly deactivating neurons during training, promoting better generalization. The last layer of this CNN model serves a specific purpose. There is also a complex but different quality. This layer provides the output of the model, which is used for classification. Notably, the number of roots in this phase equals the total number of classes in the classification function. The activation function 'softmax' used here is the standard option for multi-class classification. The 'Softmax' function converts the raw model output into a probability distribution across classes and gives the probability of each class's input data.

4.3 Ensemble Machine Learning (EML): A Collective Approach

Our study adopts an ensemble machine learning strategy to increase classification accuracy, a powerful approach that combines multiple models and exploits their collective predictive power. The primary purpose of ensemble algorithms is to obtain robust and accurate results by combining the strengths of different machine-learning models. The essence of ensemble learning lies in its ability to reduce the weakness of individual models, thus providing reliable and scalable solutions; ensemble learning tends to increase classification accuracy significantly compared to individual models. The collective insights of models can effectively reduce errors and uncertainties, leading to more accurate forecasts. Our research focuses on understanding, applying, and optimizing group strategies, emphasizing "soft voting." This method considers the prediction probabilities generated by individual models and combines them to produce aggregate predictions. Soft voting increases the ability to deal with uncertainties and complex classification tasks, making the decision-making process more effective. It measures individual models' confidence in their predictions by soft voting rather than simply counting votes. The rationale behind our choice of soft votes lies in its flexibility and effectiveness in solving complex classification problems.

4.4 Performance Considerations:

Through the research, a comprehensive evaluation of the ensemble's performance metrics has been done, which helps to understand its effectiveness in the classification tasks. The study assesses these performance metrics with accuracy, precision, recall, and F1. Together, these metrics provide a comprehensive and holistic view of the ensemble's classification competence. Accuracy reflects the percentage of correct values, precision determines the ratio of accurate optimistic predictions, recall measures the model's capability of discovering relevant instances, and the F1 score represents the average precision and recall. The particular analysis ensures analyzing the ensemble performance not through a single lens but through multiple perspectives. Secondly, the study goes further to look beyond the usual performance indicators. It examines the robustness of the ensemble in different situations. Robustness evaluation is about challenging the ensemble in different harsh conditions like noisy data, imbalanced datasets, or adversarial attacks to find out if it can give reliable predictions in the real world, where they are not ideal. Overall, such an approach enables the practitioners to assess the ensemble's resilience and flexibility.

The research also looks at the computation efficacy of ensemble methods concerning performance and robustness. It remains essential, particularly in resource-demanding applications. The paper explores how well the ensemble combines predictions from different models, using as few computational resources as possible to yield accurate results. Besides, the research traverses the complex maze of mixing many models into the ensemble. This involves choosing several models addressing compatibility issues and developing appropriate merging techniques.

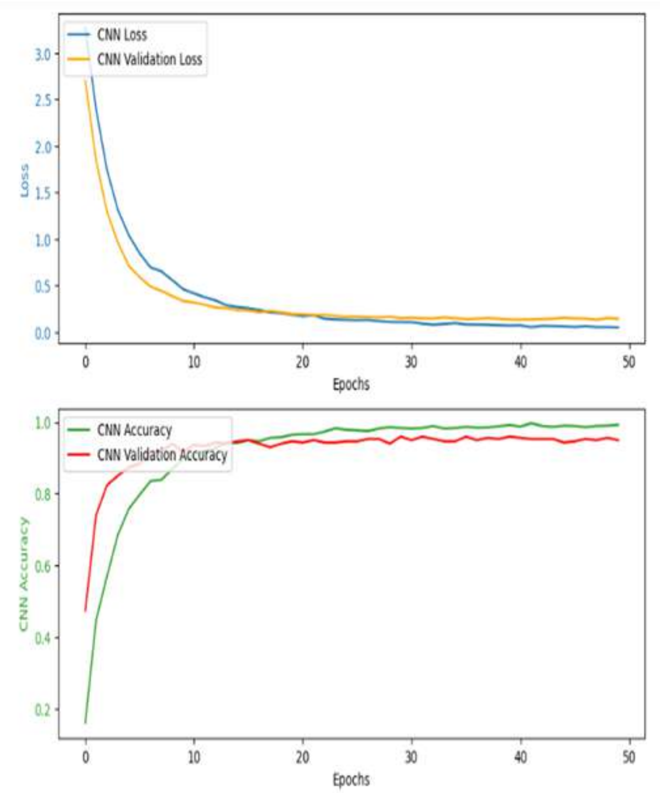


Figure 2. CNN Training Dynamics

5 Results

The dynamics of the training process of our CNN model are depicted in Figure 1. This figure presents two critical aspects of the model's learning process: accuracy and loss. The training loss or the difference between predicted and actual labels concerning the number of training epochs is depicted in the plot. The plot clearly shows that the training loss decreases steadily with the training, which indicates that the CNN model has been able to learn from the data set. It also provides a similar training and validation accuracy perspective in the lower plot. The training accuracy portrays the model's ability to correctly predict examples from the training set, which increases steadily, implying that the model fits the training data well. On the other hand, the validation accuracy measures the model's performance on a different validation set, which is a crucial assessment of the model's ability to generalize to unseen data. The graph is the CNN's learning curves plot, indicating an optimistic upward trend in validation accuracy, which indicates that the CNN can generalize and continues to improve performance.

Additionally, it provides a plot of training loss and validation. The training loss continually reduces, indicating that the CNN model is learning to reduce the error over the training dataset. Validation loss, the model performance on the validation set, also shows a decreasing trend, showing that the model can effectively generalize without overfitting. This close-up perspective on CNN training dynamics, which involves accuracy and loss, depicts the model's convergence and generalization capacity.

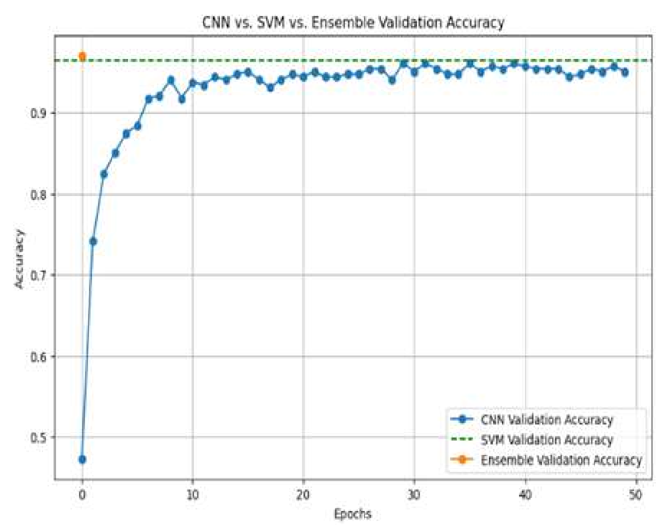
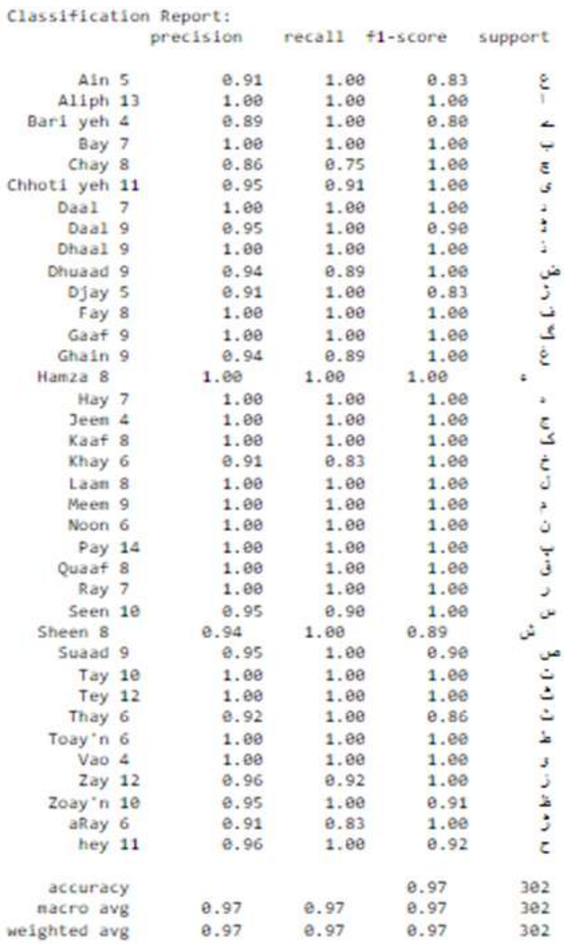


Figure 3. Accuracy Analysis

In Figure 2, we present an in-depth performance comparison of three pivotal models: our SVM-CNN ensemble model, SVM, and CNN. All these models were trained using the same dataset, and classification accuracy was a performance measure. This ensemble model is impressive, achieving 97.02% accuracy. Our result shows the tremendous consistency and the edge of our ensemble approach, which mixes the benefits of SVM and CNN from the same dataset. This shows that the ensemble model effectively classifies most images from different categories.

The classification Report summarizes the research and evaluates the performance of the USL recognition system in Figure 3. The Classification Report gives pre-



Classification Report:

	precision	recall	f1-score	support
Ain 5	0.91	1.00	0.83	ع
Aliph 13	1.00	1.00	1.00	ا
Bari yeh 4	0.89	1.00	0.80	ے
Bay 7	1.00	1.00	1.00	ب
Chay 8	0.86	0.75	1.00	چ
Chhoti yeh 11	0.95	0.91	1.00	ی
Daal 7	1.00	1.00	1.00	د
Daal 9	0.95	1.00	0.90	ڈ
Dhaal 9	1.00	1.00	1.00	ذ
Dhuaad 9	0.94	0.89	1.00	ض
Djay 5	0.91	1.00	0.83	ژ
Fay 8	1.00	1.00	1.00	ف
Gaaf 9	1.00	1.00	1.00	گ
Ghain 9	0.94	0.89	1.00	غ
Hamza 8	1.00	1.00	1.00	ء
Hay 7	1.00	1.00	1.00	ه
Jeem 4	1.00	1.00	1.00	ج
Kaaf 8	1.00	1.00	1.00	ک
Khay 6	0.91	0.83	1.00	خ
Laam 8	1.00	1.00	1.00	ل
Meem 9	1.00	1.00	1.00	م
Noon 6	1.00	1.00	1.00	ن
Pay 14	1.00	1.00	1.00	پ
Quaaf 8	1.00	1.00	1.00	ق
Ray 7	1.00	1.00	1.00	ر
Seen 10	0.95	0.90	1.00	س
Sheen 8	0.94	1.00	0.89	ش
Suaad 9	0.95	1.00	0.90	ص
Tay 10	1.00	1.00	1.00	ت
Tey 12	1.00	1.00	1.00	ٹ
Thay 6	0.92	1.00	0.86	ث
Toay'n 6	1.00	1.00	1.00	ط
Vao 4	1.00	1.00	1.00	و
Zay 12	0.96	0.92	1.00	ز
Zoay'n 10	0.95	1.00	0.91	ظ
aRay 6	0.91	0.83	1.00	ڑ
hey 11	0.96	1.00	0.92	ح
accuracy			0.97	302
macro avg	0.97	0.97	0.97	302
weighted avg	0.97	0.97	0.97	302

Figure 4. Classification Report

cision, recall, and F1-score metrics for each USL gesture. Taken together, they represent the degree of success of the model in recognizing individual signs.

It shows outstanding results and performance over different USL gestures. Notable findings include:

High Precision: Most USL gestures are displayed with high precision with values more than 0.80. This demonstrates that the model can correctly classify and separate various gestures, ensuring minimum false positives.

Strong Recall: Similarly, recall scores, which show how well the system can identify and classify relevant instances of USL gestures, are also impressive, with many gestures attaining recall scores of 1.00.

Balanced F1-Scores: The balanced F1-scores are praiseworthy and more significant than 0.90 for many gestures. One of the measures, commonly used to showcase the model's effectiveness in achieving accuracy and completeness in detecting USL gestures is popularly referred to as the F-score or simply the F1-score.

Overall Accuracy: The system's aggregate accuracy and error are only 3%. Therefore, the system correctly detects USL gestures in most situations.

Macro and Weighted Averages: The macro average also shows a balanced performance, with an F1-score of 97%. The weighted average reiterates the model's efficiency, where a weighted F1-score of 0.97 is achieved.

This implies that the USL recognition system is indeed reliable. High recall values imply the system can correctly identify different sign gestures in varying categories. The balanced F1 scores show that the model maintains both accuracy and completeness. Moreover, the macro and weighted averages also emphasize how the system is solid and effective in identifying different USL gestures.

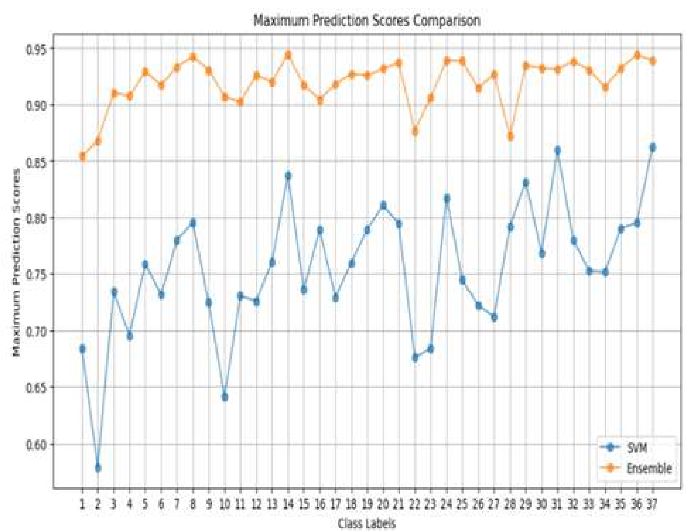


Figure 5. Prediction score Comparison

Figure 4 compares prediction scores for two models: SVM and ensemble model. This analysis aims to understand how the model performs for different classes and provides insight into the model's abilities in predicting the given dataset.

We first used the code for ensemble prediction probabilities for the test data to generate this figure. Ensemble learning is thus shown to improve predictive accuracy, the combined output of multiple classifiers being the ensemble prediction probabilities. However, the maximum prediction scores for the ensemble model ranged from 0.85 to 0.95. This means that the ensemble model continuously provides high prediction scores across the class labels, indicating that the model is reliable and better than the SVM model in predicting well.

These findings are presented in bar chart form. The bar represents the class label, from 1 to 37. The x-axis denotes the class labels, while the y-axis denotes the maximum prediction scores. The graph shows that the ensemble model always beats the SVM model as the 'Ensemble' line remains higher for all the class labels. Graphically, this illustrates the benefit of an ensemble model in this case of classification.

6 Discussion

6.1 Comparison to Existing Literature

Adopting ensemble learning as a powerful recognition strategy is new in Urdu sign language recognition. Unlike previous studies focusing on single machine learning models or specific techniques, we propose a collaborative model that combines SVM and CNN. Our ensemble model yields substantial improvement over the accuracy, where the accuracy outperforms the individual SVM and CNN models. This indicates the ability of ensemble learning to sign language recognition.

6.2 Implications

The research has both practical and theoretical implications. In practical terms, our work relates to assistive technology for the benefit of people who are hard of hearing and the deaf community. A USL recognition system development can enable this community effectively and inclusively.

As a result of examining ensemble learning as a helpful method in sign language recognition, we derive theoretical implications. We demonstrate the effectiveness of this approach and thus pave the way for more research work in the ML and sign language recognition domains. Further studies can explore ensemble methods for other sign languages to promote innovative assistive technology.

6.3 Limitations

The major limitation is that the communication process could be more fluid. The primary purpose of our system is to develop a system of converting gestures into text, which would be an essential aid for people who communicate in sign language. However, it does not provide the reverse process, for example, text transformation into sign language gestures. The unidirectional approach may not be comprehensive enough to serve the communication needs of the deaf and hard-of-hearing community, who can primarily communicate text-based messages to hearing individuals but cannot understand and respond to text-based messages they receive from others.

6.4 Future Research

Finally, we highlight several directions for future research that could improve gesture-to-text recognition for one-way communication and increase the usability of this technology for all. In the future, there is a need to develop systems that allow the deaf and hard-of-hearing community to communicate using sign language gestures and understand text-based messages. Another meaningful way is to enhance the system's recognition of contextual, non-standardized sign language gestures, considering the different styles of signing commonly used in everyday conversations. In addition, more work should be undertaken to increase accessibility and efficiency in utilizing technological resources so that they can be used by a broader population, including those in resource-poor areas. The future of gesture-to-text recognition research could be the bridge that links the gap between people who are deaf or hard of hearing and hearing worlds.

7 Conclusion

Therefore, this paper introduces a USL recognition system involving an SVM, CNN, and an ensemble model. The study explores the importance of communication tools for people who are deaf or hard of hearing and hard-of-hearing in Pakistan, where Urdu is one of the primary modes of communication.

Therefore, the research states the main objectives, such as developing an effective USL system, using machine learning techniques such as SVM and CNN, and investigating ensemble learning to increase recognition accuracy. This hypothesis tests and confirms that machine learning combined with a large dataset can be an effective USL recognition system.

The methodology consists of data collection from the "Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet" [1], normalization of the images, and HOG feature extraction. The paper presents training procedures of SVM and CNN models and an ensemble approach that uses them together for improved recognition.

We discussed hyperparameter tuning for SVM and GridSearchCV. We also discuss the architecture of the CNN model. This detailed explanation covers the approach of the ensemble model using "soft voting". Achievement, precision, recall and the F1 score of the ensemble are shown in terms of performance metrics. The study shows that the ensemble model performs better than separate SVM and CNN models, attaining an accuracy of 97.02%. These include comparing the findings with existing literature and the research implications. The paper, however, indicates a limitation as the communication process remains unidirectional and outlines future research directions to improve two-way communication systems. This research, therefore, adds to the field of assistive technology and accessibility, taking one step towards narrowing the communication gaps for the deaf and hard-of-hearing community. The assurance holds that future research will result in better gesture-to-text recognition and enhanced cohesion between deaf and hearing people.

Author Contributions

Kaneez Zainab Abbasi: Idea, Methodology, Writing-Original draft preparation **Mr. Abdul Khalique Maitlo**: Supervision **Sila Gaad, Kainat Khero**: Analysis **Kainat Khero**: Simulation work **Shazmina Khan**: Software, Validation.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] A. Imran, A. Razzaq, I. A. Baig, A. Hussain, S. Shahid, and T. U. Rehman, "Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet through Machine Learning," *Data Brief*, vol. 36, p. 107021, Jun. 2021, doi: 10.1016/j.dib.2021.107021.

[2] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 131–153, 2017, doi: 10.1007/s13042-017-0705-5.

[3] D. Bragg *et al.*, "Sign Language Recognition, Generation, and Translation," presented at the *21st Int. ACM SIGACCESS Conf. on Computers and Accessibility*, 2019.

[4] K. Parvez *et al.*, "Measuring Effectiveness of Mobile Application in Learning Basic Mathematical Concepts Using Sign Language," *Sustainability*, vol. 11, no. 11, 2019, doi: 10.3390/su11113064.

[5] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand Gesture Recognition for Sign Language Using 3DCNN," *IEEE Access*, vol. 8, pp. 79491–79509, 2020, doi: 10.1109/access.2020.2990434.

[6] J. J. Bird, A. Ekart, and D. R. Faria, "British Sign Language Recognition via Late Fusion of Computer Vision and Leap Motion with Transfer Learning to American Sign Language," *Sensors*, vol. 20, no. 18, Sep. 2020, doi: 10.3390/s20185151.

[7] S. Sharma and S. Singh, "Vision-based sign language recognition system: A Comprehensive Review," presented at the *2020 Int. Conf. on Inventive Computation Technologies (ICICT)*, 2020.

[8] U. Farooq, M. S. M. Rahim, N. Sabir, A. Hussain, and A. Abid, "Advances in machine translation for sign language: approaches, limitations, and challenges," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14357–14399, 2021, doi: 10.1007/s00521-021-06079-3.

[9] F. Shah *et al.*, "Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language," *IEEE Access*, vol. 9, pp. 67548–67558, 2021, doi: 10.1109/access.2021.3077386.

[10] M. S. Amin, S. T. H. Rizvi, and M. M. Hossain, "A Comparative Review on Applications of Different Sensors for Sign Language Recognition," *J. Imaging*, vol. 8, no. 4, Apr. 2022, doi: 10.3390/jimaging8040098.

[11] M. M. Balaha *et al.*, "A vision-based deep learning approach for independent-users Arabic sign language interpretation," *Multimed. Tools Appl.*, vol. 82, no. 5, pp. 6807–6826, 2022, doi: 10.1007/s11042-022-13423-9.

[12] D. Kothadiya *et al.*, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," *Electronics*, vol. 11, no. 11, 2022, doi: 10.3390/electronics11111780.

[13] M. S. Mirza *et al.*, "Vision-based Pakistani sign language recognition using bag-of-words and support vector machines," *Sci. Rep.*, vol. 12, no. 1, p. 21325, Dec. 2022, doi: 10.1038/s41598-022-15864-6.

[14] H. Zahid *et al.*, "Recognition of Urdu sign language: a systematic review of the machine learning classification," *PeerJ Comput. Sci.*, vol. 8, p. e883, 2022, doi: 10.7717/peerj-cs.883.

[15] N. Naz, H. Sajid, S. Ali, O. Hasan, and M. K. Ehsan, "Signgraph: An Efficient and Accurate Pose-Based Graph Convolution Approach Toward Sign Language Recognition," *IEEE Access*, vol. 11, pp. 19135–19147, 2023, doi: 10.1109/access.2023.3247761.

[16] Y. Obi *et al.*, "Sign language recognition system for communicating to people with disabilities," *Procedia Comput. Sci.*, vol. 216, pp. 13–20, 2023, doi: 10.1016/j.procs.2022.12.106.

[17] H. Zahid *et al.*, "A Computer Vision-Based System for Recognition and Classification of Urdu Sign Language Dataset for Differently Abled People Using Artificial Intelligence," *Mob. Inf. Syst.*, vol. 2023, pp. 1–17, 2023, doi: 10.1155/2023/1060135.

[18] F. Azmi and A. Novianty, "Sign Language Recognition using Principal Component Analysis and Support Vector Machine," *Int. J. Appl. Inf. Technol.*, vol. 4, no. 01, 2021, doi: 10.25124/ijait.v4i01.3015.

[19] G. Latif *et al.*, "An Automatic Arabic Sign Language Recognition System based on Deep CNN: An Assistive System for the Deaf and Hard of Hearing," *Int. J. Comput. Digit. Syst.*, vol. 9, no. 4, pp. 715–724, 2020, doi: 10.12785/ijcds/090418.

[20] H. Ahmed, S. O. Gilani, M. Jamil, Y. Ayaz, and S. I. A. Shah, "Monocular Vision-based Signer-Independent Pakistani Sign Language Recognition System using Supervised Learning," *Indian J. Sci. Technol.*, vol. 9, no. 25, 2016, doi: 10.17485/ijst/2016/v9i25/96615.