2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Attention Trinity Net and DenseNet Fusion: Revolutionizing American Sign Language Recognition for Inclusive Communication

Tanvir Shakil Joy[1], Anwar Hossain Efat [2], S. M. Mahedy Hasan[3], Nahrin Jannat [4], Mahjabin Oishe [5], Mostarina Mitu[6], A. S. M. Montashir Fahim[7]
[1,2,3,4,5,6,7]*Department of Computer Science & Engineering*
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
Emails: tnvrshakil@outlook.com,anwarhossainefat@gmail.com, mahedy@cse.ruet.ac.bd,
nahrinnipun17@gmail.com, mahjabinoishe@gmail.com,mostarinabinre@gmail.com,
montashirfahim25@gmail.com

2023 26th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-5901-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIT60459.2023.10440975

***Abstract*—American Sign Language (ASL) is a unique and essential language for the North American deaf and hard-of-hearing community, with distinct grammar and expressive hand and facial movements. In this study, we embarked on a unique journey to enhance ASL recognition, recognizing the significance of this language as a means of communication. Our approach began by consolidating two distinct datasets, amalgamating their valuable images to create a comprehensive corpus for classification. This fusion of resources provided a robust foundation for our research. A hallmark of our methodology is the incorporation of three sophisticated attention modules: channel attention, soft attention, and squeeze and excitation attention. These modules collectively form the framework known as Attention Trinity Net (ATN), a testament to our commitment to precision and performance. Furthermore, we introduced the concept of Uni Attention Network (UAN), an innovative approach that leverages DenseNet121, a powerful neural network architecture. Each attention module, channel attention, soft attention, and squeeze and excitation attention, is thoughtfully integrated into the UAN framework, enriching its ability to discern intricate ASL gestures. In a climactic juncture of our study, we seamlessly combined the outputs of the three UANs based on the ATN to form a cohesive and comprehensive prediction model. The synergy achieved through this integration resulted in an astounding accuracy rate of 99.98%, surpassing the performance of any existing ASL recognition models. Our study is a significant milestone in ASL recognition, highlighting the potential of advanced attention mechanisms and unified network structures in deep learning and computer vision. This achievement benefits the deaf and hard-of-hearing community and advances artificial intelligence and human-computer interaction.**

***Index Terms*—American Sign Language (ASL) Recognition, Attention Trinity Net (ATN), Uni Attention Network (UAN), DenseNet121.**

I. Introduction

American Sign Language (ASL) is a complete, natural language with grammar that differs from English, expressed through hand and face movements. It is the primary language for many North Americans who are deaf and serves as a key communication option for the deaf or hard-of-hearing.

Approximately 15% of American adults have some degree of hearing impairment, highlighting the importance of ASL for inclusive communication. Recent advancements in Convolutional Neural Networks (CNN) and fields like Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are significantly improving ASL recognition. These innovations help bridge communication gaps between deaf and hearing communities by enabling more efficient and real-time ASL interpretation.

A Convolutional Neural Network (CNN) model is developed for recognizing American Sign Language, employing the DenseNet-121 architecture integrated with Channel Attention, Squeeze-and-Excitation Attention, and Soft Attention mechanisms. These mechanisms aim to optimize the model's interpretability and learning capability, enabling precise and reliable ASL detection.

This exploration revolves around several pivotal questions, seeking to address the enhancement of datasets, the integration of advanced attention mechanisms, and the focus on critical features:

RQ1: Can the amalgamation of diverse datasets augment the model's interpretative proficiency and accuracy in ASL?

RQ2: What are the repercussions of implementing Channel Attention, Squeeze and Excitation Attention, and Soft Attention concurrently on the model's interpretive precision in ASL detection?

RQ3: What methodologies can be employed to focus on the most critical features, especially significant areas or regions, in ASL detection?

These questions will be addressed in the following sections of this study.

This research aims to explore the solutions, the specifics of which will be further delineated in subsequent sections. In Section III, the datasets utilized in this study are presented. The research methodology is expounded in Section IV. The findings from the experiments conducted in this study are discussed in Section V. Section VI articulates comparisons with recent studies and delineates other methodologies that have been explored. Section VII defines the limitations inherent to this study. Finally, Section VIII offers the concluding reflections and summaries of our research findings.

II. Literature Review

The proliferation of Convolutional Neural Networks (CNN) in American Sign Language (ASL) detection has spearheaded a paradigm shift in achieving heightened accuracies and refined model generalization. Pioneering studies, including those by Garcia et al. [1] and Bantupalli et al. [2], have set robust foundations with their respective accuracies of 98.00% and 98.36%, focusing on real-time detection and extensive parameter optimization. Likewise, Nguyen et al. [3] elucidated model optimization and advanced generalization techniques, yielding an accuracy of 98.36%. Furthermore, Jain et al. [4] emerged as a notable contributor, addressing critical issues of overfitting and data scarcity through innovative optimization techniques, thereby attaining an unprecedented 98.58% accuracy.

Introducing a distinctive two-tier approach, Ma et al. [5] enhanced the model's discerning capabilities regarding intricate sign details, recording an accuracy of 97.57%. Extending the scope of research, Li et al. [6] emphasized compact joints encoding for dynamic hand gesture recognition, procuring a 93.10% accuracy, providing insights into skeleton-based approaches and model compactness. Meanwhile, Kothadiya et al. [7] explored the realm of deep learning in sign language detection and recognition, achieving a 97.10% accuracy, showcasing the extensive capabilities of deep learning models in deciphering sign languages.

Several works have explored varying facets of ASL, emphasizing not only accuracy but also model interpretability and real-world applicability. Studies by Das et al. [8] and Santhalingam et al. [9] emphasized the subtleties and nuances inherent in ASL, achieving 90.00% and 93.00% accuracies respectively. In parallel, Taskiran et al. [10] emphasized the practical deployment of ASL detection models, obtaining a 98.05% accuracy, and highlighted the associated challenges and solutions in real-world scenarios.

III. Dataset Description

TABLE I
Details of Concatenated Datasets

Dataset	Classes	Images/Class	Total Images
ASL Alphabet [11]	28	3000	84,000
Synthetic ASL Alphabet [12]	27	1000	27,000
Concatenated Dataset	26	4000	104,000

The Concateneted dataset, depicted in Table I comprises a total of 104,000 images with 26 classes representing the letters A to Z. Each class is balanced, containing 4000 images.

IV. Research Methodology

To construct a comprehensive dataset, two distinct datasets were concatenated to form a cohesive and enriched collection, incorporating the diverse information from each. Following this concatenation, non-shared classes between the original datasets were meticulously identified and removed, ensuring uniformity and coherence in the classes within the consolidated dataset. This process was crucial in eliminating disparities and inconsistencies, thereby enhancing the accuracy and reliability of subsequent analyses.

The project involved integrating a pre-trained Convolutional Neural Network, DenseNet-121, chosen for its dense connectivity and efficient architecture. To augment its feature interpretability and focus on key regions, three distinct attention mechanisms were multiplicatively incorporated, creating the Attention Trinity Net (ATN). Initially, a Channel Attention mechanism was applied to selectively model interdependencies between channels, emphasizing discriminative features. This was followed by the integration of a Squeeze-and-Excitation (SE) block, which adaptively recalibrates channel-wise feature responses, highlighting the most informative channels. Lastly, a Soft Attention mechanism assigned attention scores to each spatial location in the feature map, enabling the model to focus on specific regions of the input image. The combination of DenseNet-121 with these attention mechanisms aimed to enhance the ATN model's ability to capture complex patterns and intricate details, improving performance and robustness in various tasks.

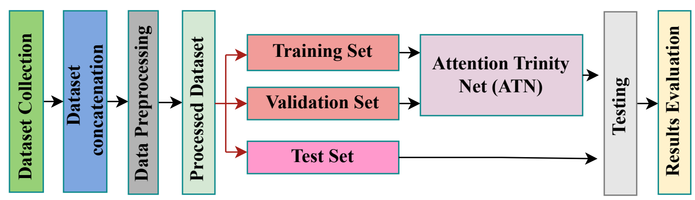

Fig. 1. Sequential Flow-Diagram of The Study.

A. *Preprocessing*

Preprocessing always plays a vital role in data and feature selection [13], [14]. In the course of developing a comprehensive dataset, two distinct datasets were integrated. The first dataset comprised 28 classes, whereas the second included 27 classes. To establish coherence and maintain relevance to the overarching research objectives, a meticulous selection process was undertaken to retain only the 26 classes common to both datasets, effectively eliminating the non-concurring classes.

Subsequently, the consolidated dataset was partitioned into training, testing, and validation subsets, in accordance with established best practices for machine learning endeavors. The distribution ratio for the split was set at 70% for training, 15% for testing, and 15% for validation. This strategic allocation ensures an optimal balance, allowing for robust model training, effective evaluation, and validation to assess the model's generalization capability on unseen data.

B. *Proposed Attention Trinity Net (ATN) Architecture*

A DenseNet 121 (DN 121) Convolutional Neural Network (CNN) model is employed for the task, meticulously structured with an initial four convolutional layers, each integrating 128 filters, and followed by another set of four convolutional layers, each with 256 filters. To stabilize the activations and address the internal covariate shift, Batch Normalization is applied after every convolutional layer.

To describe the process more formally, we can start with three distinct attention mechanisms. The outputs from these mechanisms are concatenated, forming a cohesive structure. This unified output is then subjected to a flattening layer, reshaping the input into a one-dimensional array to ensure compatibility with the subsequent dense layers. After the flattening layer, the model introduces densely connected neural network layers, or 'Dense' layers. The first Dense layer incorporates 1024 units, succeeded by another Dense layer consisting of 512 units. The final Dense layer is composed of 26 units, corresponding with the required output dimensionality.

The graphical representation of the architecture is depicted in Figure 2.

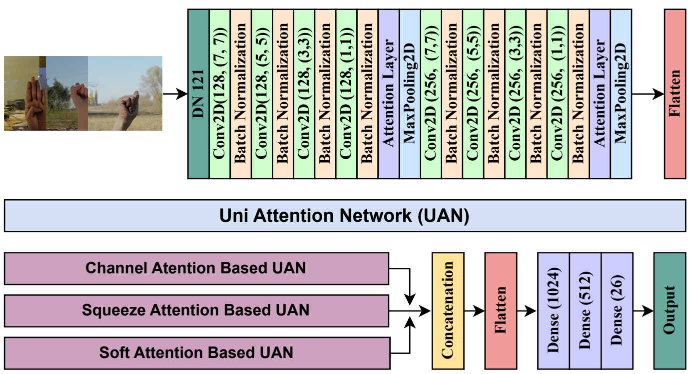

Fig. 2. Attention Trinity Net (ATN) Architecture.

This cascading configuration of attention mechanisms and densely connected layers henceforth facilitates the effective transformation and extraction of relevant features from the input, eventually leading to the generation of the final output.

The convolutional layers were incorporated with various kernel sizes, including 7x7, 5x5, 3x3, and 1x1, to diversify the receptive fields and extract hierarchical features from the input. A MaxPooling2D layer was interposed between the two sets of convolutional layers with 128 and 256 filters to reduce the spatial dimensions and intensify the representative power of the feature maps. Subsequent to the convolutional components, an Attention Layer was integrated to weigh and prioritize the pertinent features adaptively. Following the Attention Layer, another MaxPooling 2D layer was applied to further condense the spatial dimensions of the feature maps. Finally, a Flatten operation was performed to convert the 2D matrix data to a vector, making it suitable for the ensuing fully connected layers or output layers. This intricate design aspired to ensure a robust and effective feature extraction and representation capability within the developed model.

C. *Attention Trinity Net (ATN)*

The ATN model employs attention modules to selectively emphasize essential features within the input while minimizing the influence of less relevant ones. This model is characterized by the incorporation of three distinctive attention modules to reinforce this emphasis.

1) Channel Attention (CA): The Channel Attention (CA) module refines feature maps through the computation of channel-wise attention weights, derived from both the mean and standard deviation [15]. These weights are subsequently applied to the input feature maps to enhance essential features. The mathematical formulation is:

$$w_c = \sigma(W_2 \delta(W_1 x)) \tag{1}$$

In this representation, x denotes the input feature maps of dimensions $C \times H \times W$, and W_1 and W_2 represent learnable weight matrices. Additionally, δ signifies the ReLU activation function, σ denotes the sigmoid activation function, and w_c corresponds to the determined attention weights for each channel.

$$y_c = w_c \odot x \tag{2}$$

Here, $\odot$ represents element-wise multiplication, and y_c stands for the resultant, feature-enhanced maps.

2) Squeeze and Excitation Attention (SEA): The SEA module performs two primary operations: a squeeze operation that diminishes the spatial dimension of the input feature maps, and an excitation operation which determines channel-wise attention weights to amplify pivotal features.

$$z = \text{GlobalAveragePooling}(x) \tag{3}$$

$$s = \text{ReLU}(W_2 \sigma(W_1 z)) \tag{4}$$

$$y = s \odot x \tag{5}$$

In this context, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are learnable weight matrices. The symbol r represents the reduction ratio, and $\odot$ indicates element-wise multiplication.

3) Soft Attention (SA): Soft Attention (SA) allocates weights to distinct elements within the input to discern their respective significance, thus allowing the model to focus selectively on certain input parts.

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)} \tag{6}$$

Here, α_i symbolizes the attention weight of the i-th element in the input sequence. T is the length of the sequence, and e_i is the associated scalar score for the i-th element.

D. *Justification of Our Proposed Architecture*

The presented work integrates multiple attention models: Channel Attention, Squeeze and Excitation, and Soft Attention, to examine their effectiveness in resolving research problems. Batch Normalization is crucial in stabilizing the neural network's training and optimizing the process.

The Attention Trinity Net, merging three attention modules, maximizes each mechanism's potential. This integration is key to obtaining synergistic advantages and enhancing model performance.

Employing filters of diverse sizes is vital in apprehending features at multiple scales, enriching the model's proficiency in deciphering complex patterns. The fusion of these attention-centric networks forms a comprehensive strategy, essential for constructing a resilient and inclusive model.

In conclusion, combining varied attention models, batch normalization, and different filter sizes refines the model's resilience and adaptability to new data, highlighting its relevance for the conference.

E. Performance Evaluation Measures

To evaluate the efficacy of the models, a variety of metrics are utilized, encompassing accuracy, precision, recall (sensitivity), F1-score, specificity, and Receiver Operating Characteristic Area Under Curve (ROC-AUC). These metrics are represented mathematically as follows [16]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{11}$$

F. Experimental Setup

The whole architecture is run on a Kaggle notebook, leveraging the NVIDIA TESLA P100 GPU. Distinct lesion images with an input size of $(224, 224, 3)$ were obtained, and subsequently, 15% of the dataset was reserved for validation, and another 15% for testing, with the remaining images utilized for the training phase. The training of the models was performed over 35 epochs with a batch size of 16, utilizing the Adam optimizer with a learning rate set at 0.001. The loss function applied was Categorical Cross-Entropy, and a Reduce on Plateau strategy was adopted for early stopping, with a patience level of 25.

V. Experimental Results Analysis

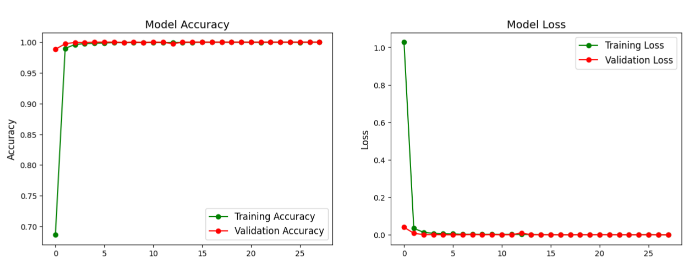

Fig. 3. Model Performance: Accuracy and Loss Curves

From the analysis of the accuracy and loss curves shown in Figure 3, it is evident that the model attains maximal accuracy after processing a minimal number of epochs, subsequently stabilizing for the remaining iterations. To mitigate the risk of overfitting, a patience parameter of 25 was utilized in the training process.

In the evaluation of the classification model encompassing 26 distinct classes, every class manifested impeccable accuracy, registering at 100%, with the exception of the 21st and 25th classes. These two classes demonstrated a marginal deviation, delivering an accuracy of 99.83%.

Moreover, the model exhibited flawless precision across all classes, scoring 100%, barring the 20th class. The precision for this class was recorded at 99.67%. This marginal diminution in precision for the 20th class and accuracy for the 21st and 25th classes serve as the sole aberrations in an otherwise exemplary model performance.

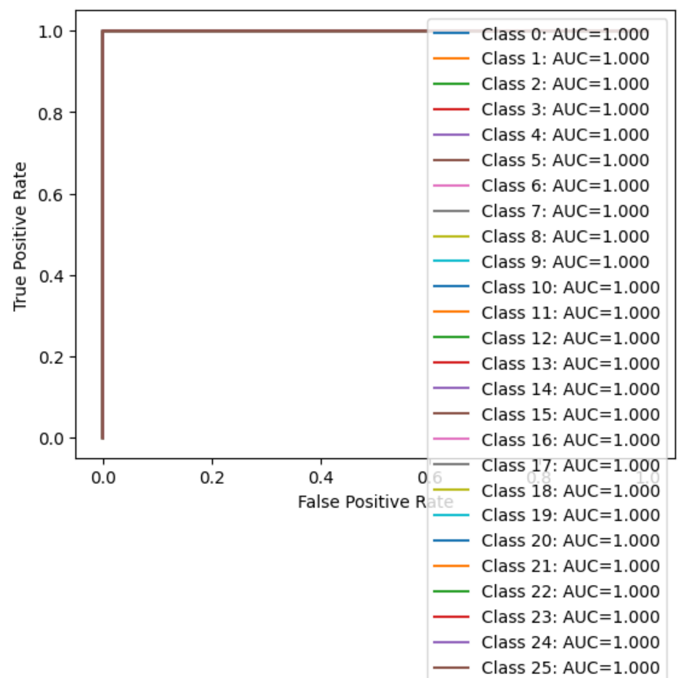

Fig. 4. Receiver Operating Characteristic (ROC) curve

Based on the Figure 4 of the analysis of the Receiver Operating Characteristic (ROC) curve, it is observed that the convergence rate is notably high across all classes which is actually the ideal scenario based on [17], [18]. Moreover, the Area Under the Curve (AUC) is calculated to be 1 for every class, indicating perfect discriminatory power and optimal model performance for each respective class. This implies that the model demonstrates exemplary predictive capabilities and is able to distinguish effectively between the different classes in the dataset, thereby ensuring reliability and robustness in its predictive applications.

The Confusion Matrix in Figure 5, which is employed to evaluate the classification accuracy and to identify the patterns of misclassification, illustrates that classes V and Z are the only ones exhibiting misclassification errors. Specifically, each of these classes has one instance where it is misclassified as the other; that is, one instance of class V is misclassified as class Z, and reciprocally, one instance of class Z is misclassified

TABLE II
PERFORMANCE EVALUATION OF THE PROPOSED ARCHITECTURE

Method	Accuracy	Precision	Recall	Specificity	F1 Score	MCC	Kappa
DenseNet-121	99.98	99.98	99.98	99.99	99.98	99.99	99.99

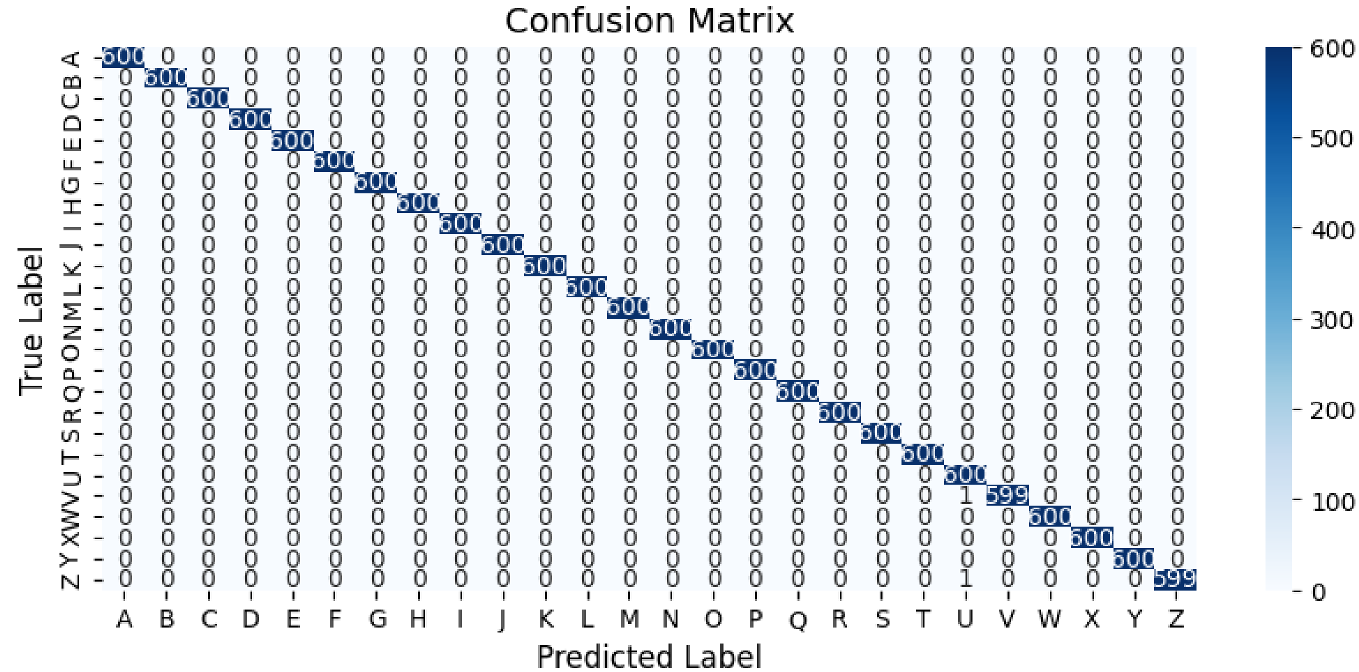

Fig. 5. Confusion Matrix of the Proposed Architecture.

as class V. All the other classes (A through U, W, and X to Y) are classified with absolute precision, displaying no misclassifications in the matrix, thereby showcasing an overall high accuracy in the classification model, with minor discrepancies confined to classes V and Z.

Answer to RQ1:

To enhance the model's interpretative proficiency and accuracy in ASL, the amalgamation of diverse datasets can be implemented, focusing on integrating varied sample types to enrich the dataset's comprehensiveness. This amalgamation can lead to improved model performance by providing a broader spectrum of data points for the model to learn diverse features, especially in ASL detection. If the model is trained with datasets that lack diversity, it may not effectively interpret the wide range of signs in ASL. Therefore, it is recommended to amalgamate diverse datasets for training but to be cautious during validation and testing to avoid overfitting and ensure model generalization.

Answer to RQ2:

The implementation of Channel Attention (CA), Squeeze and Excitation Attention (SEA), and Soft Attention (SA) concurrently can have profound repercussions on the model's interpretive precision in ASL detection. When these attention mechanisms are embedded within the model architecture, they allow the model to focus on different aspects of the input, ensuring that the model grasps the relevant information and nuanced features of ASL. Analyzing the outcomes, it is apparent that the simultaneous incorporation of these attention mechanisms can significantly elevate the model's performance in detecting and interpreting ASL compared to models without these attention layers, by allocating varied attention to diverse features in the input.

Answer to RQ3:

In ASL detection, it's crucial to focus on significant areas or regions, which can be achieved by incorporating advanced attention mechanisms and feature extraction techniques in the model architecture. These mechanisms allow the model to prioritize crucial features and, by capturing both shallow and deep features, ensure no critical information is missed. This approach significantly enhances the model's accuracy and reliability in ASL detection by ensuring it effectively captures and emphasizes the most important features.

By focusing on these modified approaches and implementations, models can be refined to ensure optimized interpretative proficiency and heightened accuracy in ASL detection, making them more reliable and robust for practical applications.

VI. DISCUSSION AND EXTENDED COMPARISON

In conclusion, the proposed model, with an accuracy of 99.98%, outperforms all the listed models in terms of accuracy

TABLE III
COMPARISON OF DIFFERENT MODELS

Authors	Accuracy	Precision	Recall	F1-score	Specificity	MCC	Kappa
[1]	98.00%	-	-	-	-	-	-
[2]	98.36%	-	-	-	-	-	-
[4]	97.57%	-	-	-	-	-	-
[5]	97.57%	94.48%	94.56%	94.52%	-	-	-
[8]	90.00%	-	-	-	-	-	-
[9]	93.00%	-	-	-	-	-	-
[10]	98.05%	-	-	-	-	-	-
[3]	98.36%	-	-	-	-	-	-
[6]	93.10%	-	-	-	-	-	-
[7]	98.36%	97.00%	97.00%	97.00%	-	-	-
Ours	**99.98%**	**99.98%**	**99.98%**	**99.98%**	**99.99%**	**99.99**	**99.99**

on American Sign Language detection, underscoring the potential for advancements in real-time translation, educational tools, and enhanced accessibility in public spaces.

VII. THREATS TO VALIDITY

The limitations of this research are primarily anchored in its exclusive focus on American Sign Language (ASL), rendering the findings non-generalizable to other global sign languages, each with its unique grammatical structure and lexicon. The variability inherent in the data collection process, such as differing lighting, backgrounds, and signer demographics, poses substantial threats to the external validity of the study. Additionally, the evolving nature of sign languages and limitations in the experimental setup and chosen evaluation metrics may impact the robustness and long-term applicability of the developed CNN model.

VIII. CONCLUSION AND FUTURE WORK

Our research represents a significant advancement in American Sign Language (ASL) recognition, a crucial communication medium for the deaf and hard-of-hearing community in North America. ASL, distinct in its grammatical structure and reliance on complex hand and facial movements, differs fundamentally from spoken languages like English. We merged two datasets and utilized a novel method, the Attention Trinity Net (ATN), incorporating three attention modules—channel attention, soft attention, and squeeze and excitation attention. Each module played a key role in enhancing our model's performance.

Furthering our innovation, we developed the Uni Attention Network (UAN), integrated with DenseNet121, to significantly boost ASL recognition. This approach, involving the concatenation of three UANs based on ATN, achieved a groundbreaking 99.98% accuracy, surpassing existing models.

This research marks a leap in ASL recognition and highlights the potential of sophisticated attention mechanisms and unified network architectures in deep learning and computer vision. It contributes to both the deaf and hard-of-hearing community and the broader fields of artificial intelligence and human-computer interaction. Our ongoing efforts aim to extend these technologies to recognize other sign languages, bridging communication gaps across diverse communities.

REFERENCES

[1] B. Garcia and S. A. Viesca, "Real-time american sign language recognition with convolutional neural networks," *Convolutional Neural Networks for Visual Recognition*, vol. 2, no. 225-232, p. 8, 2016.

[2] K. Bantupalli and Y. Xie, "American sign language recognition using deep learning and computer vision," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 4896–4899.

[3] H. B. Nguyen and H. N. Do, "Deep learning for american sign language fingerspelling recognition system," in *2019 26th International Conference on Telecommunications (ICT)*. IEEE, 2019, pp. 314–318.

[4] V. Jain, A. Jain, A. Chauhan, S. S. Kotla, and A. Gautam, "American sign language recognition using support vector machine and convolutional neural network," *International Journal of Information Technology*, vol. 13, pp. 1193–1200, 2021.

[5] Y. Ma, T. Xu, and K. Kim, "Two-stream mixed convolutional neural network for american sign language recognition," *Sensors*, vol. 22, no. 16, p. 5959, 2022.

[6] Y. Li, D. Ma, Y. Yu, G. Wei, and Y. Zhou, "Compact joints encoding for skeleton-based dynamic hand gesture recognition," *Computers & Graphics*, vol. 97, pp. 191–199, 2021.

[7] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, p. 1780, 2022.

[8] A. Das, S. Gawde, K. Suratwala, and D. Kalbande, "Sign language recognition using deep learning on custom processed static gesture images," in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*. IEEE, 2018, pp. 1–6.

[9] P. S. Santhalingam, P. Pathak, H. Rangwala, J. Košecká *et al.*, "Finehand: Learning hand shapes for american sign language recognition," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020, pp. 700–707.

[10] M. Taskiran, M. Killioglu, and N. Kahraman, "A real-time system for recognition of american sign language by using deep learning," in *2018 41st international conference on telecommunications and signal processing (TSP)*. IEEE, 2018, pp. 1–5.

[11] "Asl alphabet," https://www.kaggle.com/datasets/grassknoted/asl-alphabet, (Accessed on 09/30/2023).

[12] "Synthetic asl alphabet," https://www.kaggle.com/datasets/lexset/synthetic-asl-alphabet, (Accessed on 09/30/2023).

[13] S. Datta, S. M. Hasan, M. F. Faruk, A. Y. Srizon, M. R. Hossain, A. H. Efat, and M. Mitu, "Improved diabetes prediction with reduced feature sets: Evaluating feature selection techniques in machine learning," in *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2023, pp. 104–108.

[14] S. Datta, S. M. Hasan, M. Mitu, M. F. Taraque, N. Jannat, and A. H. Efat, "Hyperparameter-tuned machine learning models for complex medical datasets classification," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2023, pp. 1–6.

[15] N. Jannat, S. M. Hasan, A. Y. Srizon, M. Oishe, A. H. Efat, M. F. Taraque, M. Mitu, and M. F. Faruk, "Efficient detection of crop leaf diseases: A lightweight convolutional neural network approach for enhanced agricultural productivity," in *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2023, pp. 99–103.

[16] M. Mitu, S. M. Hasan, A. H. Efat, M. F. Taraque, N. Jannat, and M. Oishe, "An explainable machine learning framework for multiple medical datasets classification," in *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*. IEEE, 2023, pp. 1–6.

[17] A. H. Efat, S. M. Hasant, N. Jannat, M. Mitu, M. F. Taraque, S. Ferdous, and S. Datta, "Inquisition of the support vector machine classifier in association with hyper-parameter tuning: A disease prognostication model," in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. IEEE, 2022, pp. 131–134.

[18] M. F. Taraque, S. M. Hasan, N. Jannat, A. H. Efat, M. Mitu, S. Datta, and S. Ferdous, "Early stage prediction of autism spectrum disorder: Analyzing different hyperparameter tuned machine learning classifier," in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. IEEE, 2022, pp. 1–4.