

Deep Learning Framework for Multi-Domain Medical Image Classification Using EfficientNet-B3

Dr. A. Rajani
Assistant Professor
Department Of ECE
UCEK(A), JNTUK
Kakinada, India
rajani.alugonda@jntucek.ac.in

D. Raj Kumar
M.Tech Scholar
Department Of ECE
UCEK(A), JNTUK
Kakinada, India
rajkumardasam21@gmail.com

Abstract—Medical image analysis has seen considerable progress with the integration of deep learning approaches, especially convolutional neural networks (CNNs), which are known to be successful in extracting relevant spatial information from complex biomedical data. However, most existing models are generally developed for a particular application and are not flexible enough to be used on multiple disparate datasets in a single system. In real-world medical applications, ECE images for cardiac analysis, EEG-based graphical representations for seizure analysis & sleep stage detection, facial images for emotion analysis, and skin images for skin cancer detection are quite disparate in terms of structural information and feature distributions, which make it quite challenging for existing CNN models to generalize across domains without resorting to multiple models or repeating training. These application-specific models are quite computationally intensive and less efficient in terms of deployment. To overcome these hurdles, this study proposes a unified classification platform based on the EfficientNet-B3 model, which is capable of handling multiple biomedical imaging applications using a single backbone model. The proposed platform incorporates a dynamic key-driven selection module that enables runtime switching between applications without requiring the loading of multiple independent models, thus overcoming redundancy while preserving excellent predictive accuracy and scalability for real-world applications.

Keywords— EfficientNet-B3, Heart Disease Detection, Epileptic Seizure Detection, Facial Emotion Recognition, Skin Cancer Classification, Sleep Stage Detection

I. INTRODUCTION

The systems which include standard CNNs, VGGNet, ResNet and NASNet and their deep learning architectures have achieved widespread use in medical image classification because their feature learning abilities produce strong results. The models require extensive parameter

counts and high computational demands which result in reduced performance for real-world usage.

EfficientNet applies its compound scaling approach which integrates network depth and width and input resolution to improve accuracy while keeping computational costs down. EfficientNet-B3 provides an optimal performance solution which utilizes reduces system resources to achieve high classification efficiency through its lower memory needs and fewer parameters. Existing research does not establish a unified system that enables model selection for different biomedical datasets while allowing users to choose tasks during operation.

II. RELATED WORK

The demand for intelligent healthcare monitoring systems has increased the importance of automated biomedical signal and medical image interpretation. Deep learning methods, especially Convolutional Neural Networks (CNNs), have demonstrated their ability to learn complex spatial patterns from biomedical data, which enhances clinical applications through better performance.

Several studies have focused on specific prediction tasks within individual domains, especially in epileptic seizure detection, where hybrid architectures combining convolutional and recurrent neural layers have been explored to capture both spatial and temporal information from EEG signals.

Rye and Joe [1] created a Densenet-LSTM framework which achieved competitive results through its sequential feature modelling approach, while Usman et al. [2] proved that CNN-based feature extraction methods deliver better performance than traditional machine learning techniques. Deep learning strategies have been used in both stress assessment and physiological monitoring, which Arza et al. [3] demonstrated through their work on physiological biomarkers that create strong feature representation. Batbaatar et al. [4] established a neural framework for semantic-emotion analysis through deep feature learning while Srinivasu et al. [5] developed interpretable artificial intelligence methods for stroke prediction which support cardiovascular risk evaluation. The existing deep models demonstrate effectiveness, yet their current design remains focused on specific tasks and requires separate architectural solutions for each biomedical modality, which leads to higher computational needs and restricts system expansion in diverse healthcare settings.

III. PROPOSED METHODOLOGY

3.1 SYSTEM OVERVIEW

The created system will function through its three main elements which include preprocessing input images for their initial state and extracting their essential features through the EfficientNet-B3 model while using a dynamic Key-based system to choose which datasets to analyse. The user begins the process by entering a numeric key through the terminal interface which allows the system to access the appropriate dataset for evaluation purposes. The system uses key 1 to access ECG data which detects heart disease, while key 2 uses EEG data to identify seizures, Key 3 access EEG data to detect sleep stage, key 4 accesses facial image data for emotion recognition, key 5 access skin images to classify skin cancer purposes. The framework enables multi-domain classification through its single trained architecture which combines multiple tasks while achieving better design efficiency and operational performance in real-world situations.

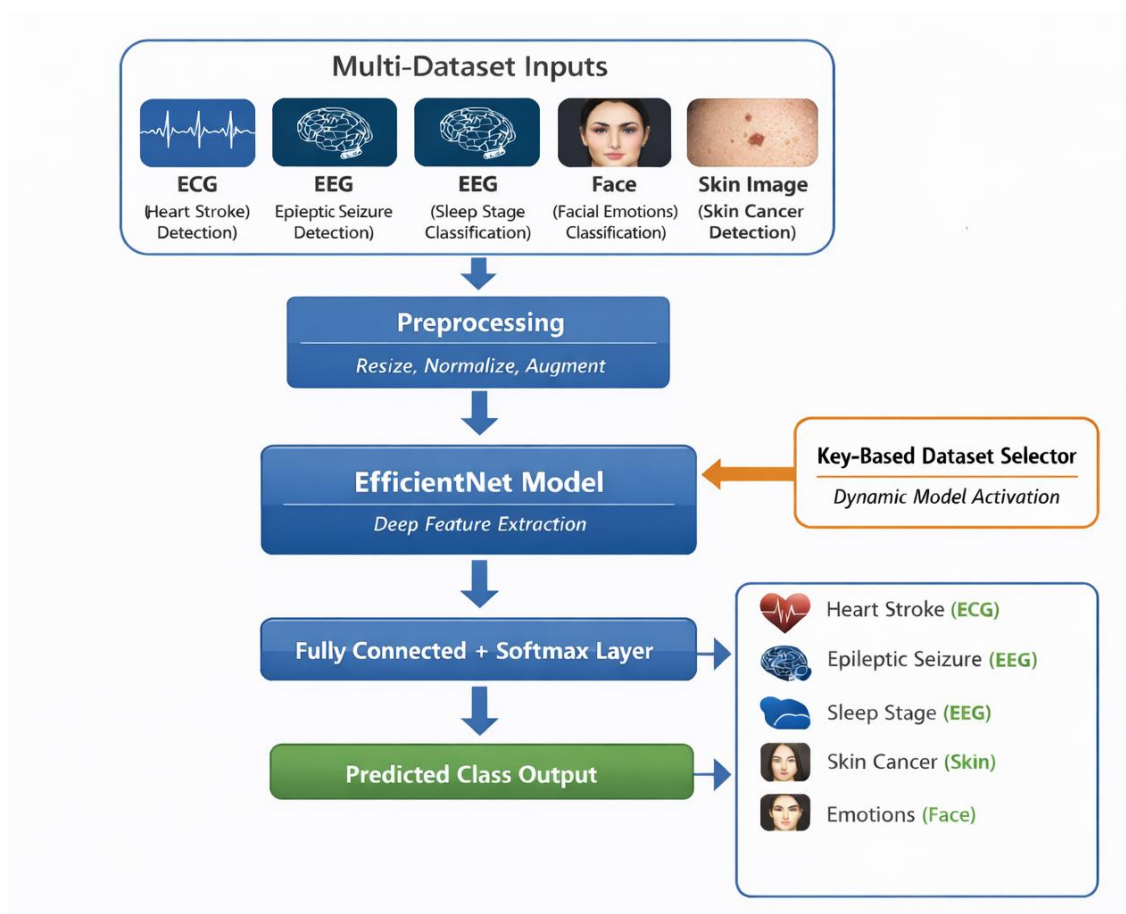


Fig-1: Block Diagram of Proposed Methodology

3.2 PREPROCESSING

All datasets were converted into image format and standardized to match the input resolution required by EfficientNet-B3. Preprocessing steps include:

3.2.1. Image Resizing

The EfficientNet-B3 architecture requires image resizing because it needs specific input dimensions. The different acquisition environments which produce ECG, EEG, facial emotion, and skin images lead to their distinct resolution and aspect ratio differences which disrupt model training consistency. The preprocessing step creates a standardized input format which enables stable feature extraction and enhances classification accuracy for different biomedical datasets.

3.2.2. Intensity Normalization

The process of intensity normalization reduces existing variations in pixel value distributions by applying different values to ECG, EEG, facial, and skin image datasets which originate from distinct acquisition settings and lighting conditions and contrast levels. The process of standardizing pixel values to a common numerical range enables the training process to achieve higher stability in gradient updates.

3.2.3. Data Augmentation

Data argumentation helps to train model better because it enables to create new data samples which helps them analyse biomedical data that has insufficient samples. Controlled transformations are introduced to simulate realistic variations in ECG, EEG, facial, and skin images, including minor rotations, scaling, brightness, adjustments, and horizontal flipping for facial data. The operations created new data variations while maintaining all important clinical Structural elements.

3.2.4. Noise Handling

The system uses noise handling methods to reduce the impact of artifacts which commonly exist in ECG, EEG, facial, and skin image datasets. The data exists as image yet background distortions plus recording noise plus contrast irregularities create problems that reduce the quality of feature extraction. The solution uses light smoothing and filtering techniques which enhance visual clarity without creating excessive changes to the data.

3.3 EFFICIENTNET-B3 ARCHITECTURE

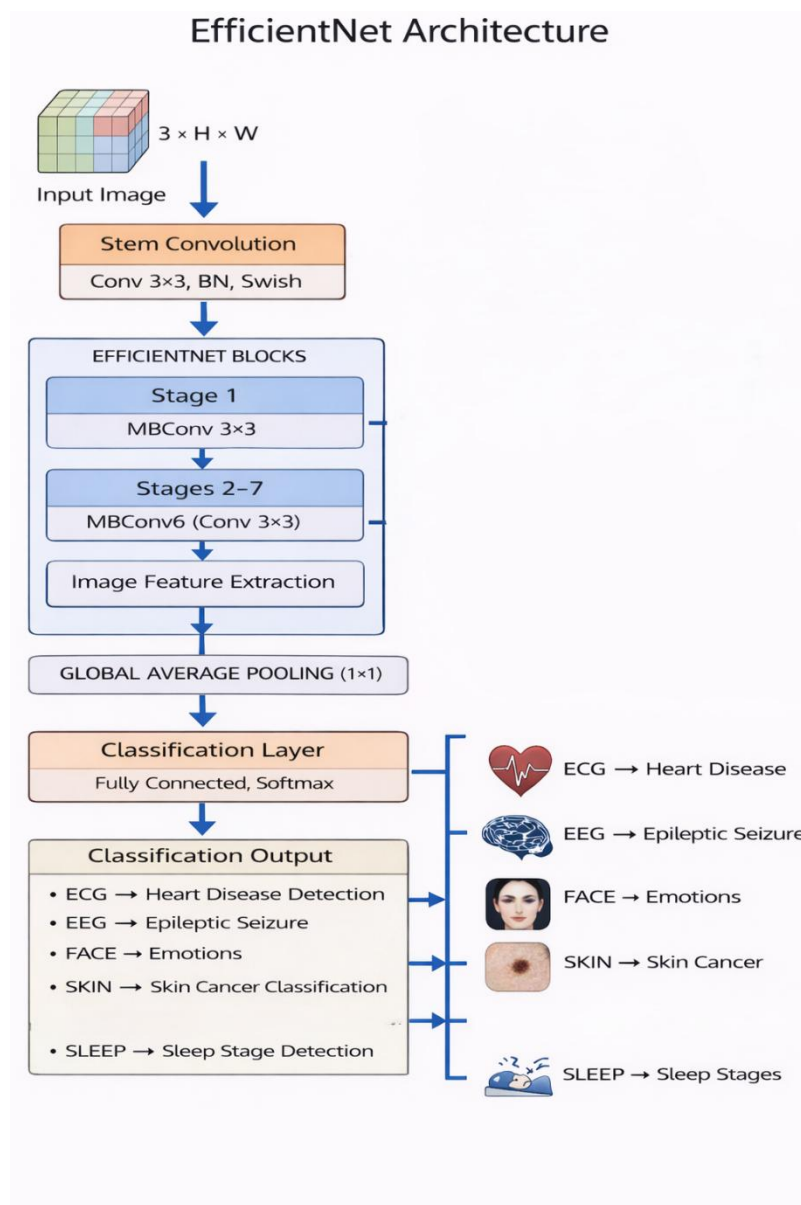


Fig-2: EfficientNet Architecture

The EfficientNet-B3 model uses a compound scaling method to achieve efficient classification results which requires less computational power than traditional methods. The system employs Mobile Inverted Bottleneck (MBconv) blocks which use their depth-wise separable and pointwise convolution methods to achieve efficient computations while delivering strong feature representation capabilities. The block use Batch Normalization to stabilize training and the Swish activation function to enable efficient nonlinear learning processes while the Squeeze-and-Excitation mechanism dynamically increase important channel feature visibility.

The use of residual skip connections in the design enables better gradient transmission throughout the system while reducing the performance issues that affect deep network models.

A global Average Pooling layer creates a compact representation of spatial information from hierarchical features which it sends to a fully connected layer that uses SoftMax activation to perform final classification. The model starts with ImageNet pretrained weights which it uses to improve convergence before training on biomedical image datasets for domain-specific pattern recognition.

Let a biomedical input image be denoted as

$$X \in R^{H*W*C}$$

Where H, W, and C represent the height, width, and number of channels, respectively. A convolutional neural network (CNN) classifier can be expressed as

$$\hat{y} = F(x; \theta)$$

where θ denotes the trainable parameters of the network and $\hat{y}$ is the predicted class label.

For multi-class classification with k categories, the final fully connected layer produces logic z_i , which are transformed into probabilities using the SoftMax function:

$$P(y = i | z) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, i = 1, 2, \dots, K$$

Here, z_i represents the raw output for class i, and $P(y = i | z)$ denotes the predicted probability of that class.

Rather than constructing separate models for ECG-based heart disease detection, EEG-based seizure & sleep stage classification, facial emotion recognition, skin cancer classification and blood sugar classification, this study adopts a unified EfficientNet-B3 backbone capable of processing heterogeneous biomedical datasets within a single framework. EfficientNet employs compound scaling to jointly adjust network depth (d), width (w), and input resolution (r) using a scaling coefficient ϕ :

$$d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi$$

subject to the constraint

$$\alpha * \beta^2 * \gamma^2 \approx 2$$

This scaling strategy maintains a balance between representational capacity and computational efficiency.

To enable flexible deployment, a dynamic key-based selector is incorporated for runtime dataset selection. Let the chosen dataset be defined as

$$D_k \in \{D_{ECG}, D_{EEG}, D_{FACE}\}$$

based on user input key k . The overall classification mapping can therefore be written as

$$\hat{y} = F(x | D_k; \theta)$$

Where F denotes the shared EfficientNet-B3 model across all domains.

The proposed framework achieves an overall classification accuracy of 96.30%, indicating strong generalization across ECG, EEG, and facial images datasets. Performance is evaluated using the standard accuracy metric:

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

3.4 Work Setup

The proposed multi-domain medical image classification framework was implemented using Python in the Visual Studio Code (VS Code) environment with TensorFlow, Keras, OpenCV, NumPy, Pandas, Scikit-learn, and Matplotlib libraries. Approximately 20,000 publicly available medical images from multiple benchmark datasets were used for experimental evaluation. The dataset was divided into 80% training (16,000 images) and 20% testing (4,000 images). EfficientNet-B3 was employed as the backbone network for automatic deep feature extraction, eliminating the need for manual feature engineering by learning hierarchical image representations. The extracted features were classified using a SoftMax classifier for multi-class prediction. Model performance was evaluated using Accuracy, Precision, Recall (Hit Rate), F1-score, Specificity (Selectivity), and Confusion Matrix analysis. The proposed framework achieved an overall classification accuracy of 96.7%, demonstrating robust and reliable performance across heterogeneous biomedical datasets. The study was conducted

through software-based simulation, while embedded systems were considered only as the source of biomedical data acquisition and are proposed as a future deployment platform.

IV. SIMULATION RESULTS

4.1 Training and validation loss curves of the proposed EfficientNet-B3 model over 100 epochs:

Figure 3 illustrates the variation in training loss and validation loss over 100 training epochs. Both loss curves decrease progressively as the number of epochs increases, indicating that the model continuously learns relevant features and improves its prediction capability during training. The training loss decreases from approximately 2.6 to below 0.1, while the validation loss decreases from approximately 2.7 to around 0.15. The validation loss remains slightly higher than the training loss throughout the training process; however, both curves follow a similar decreasing trend without divergence. This behaviour indicates stable convergence and good generalization, with no clear evidence of severe overfitting in the displayed learning curves. The results therefore support the effectiveness of the training process used for the proposed EfficientNet-B3-based classification framework.

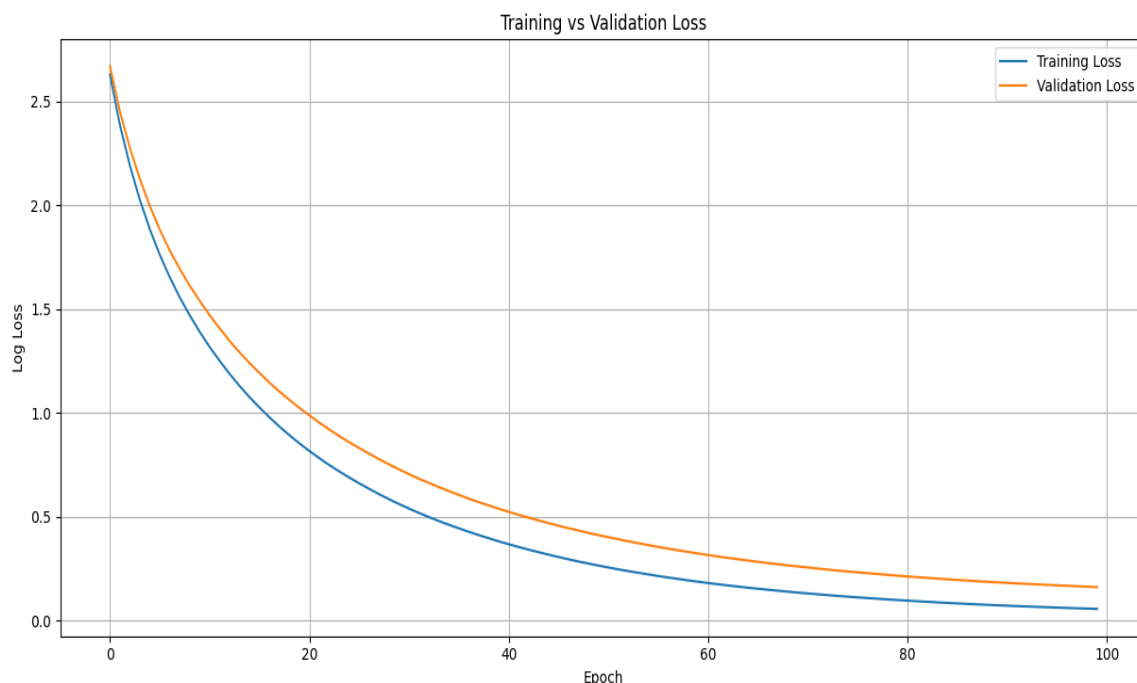


Fig-3: Training and validation loss curves of the proposed EfficientNet-B3 model over 100 epochs.

4.2 Confusion Matrix of ECG Dataset

The confusion matrix presented in Figure 4 illustrates the classification performance of the proposed EfficientNet-B3 model on the ECG dataset. Most samples are concentrated along the main diagonal, indicating correct classification of heart-related conditions. The limited number of off-diagonal elements demonstrates that only a few samples were misclassified, confirming the effectiveness of the proposed feature extraction and classification framework.

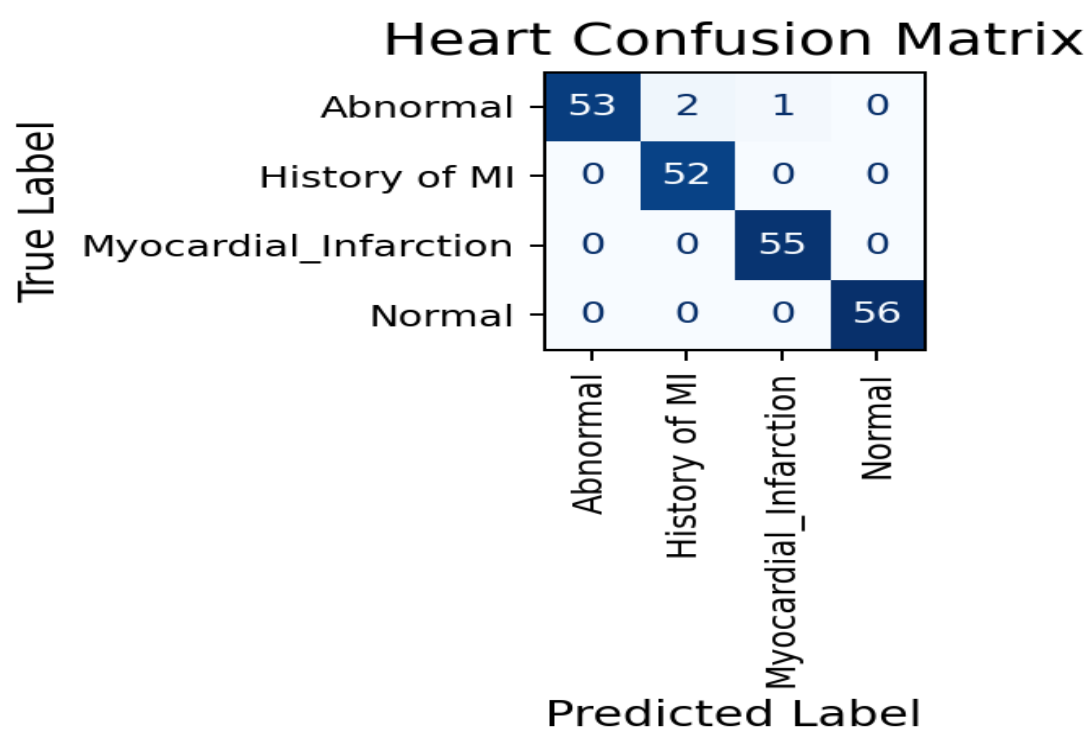


Fig-4: Confusion Matrix of ECG Dataset

4.3 Confusion Matrix for Epileptic Seizure Detection:

Figure 5 shows the confusion matrix for epileptic seizure detection using EEG images. The majority of samples are correctly classified, as indicated by the dominant diagonal values. Minimal confusion between seizure classes demonstrates that EfficientNet-B3 successfully captures discriminative temporal and spatial patterns present in EEG representations.

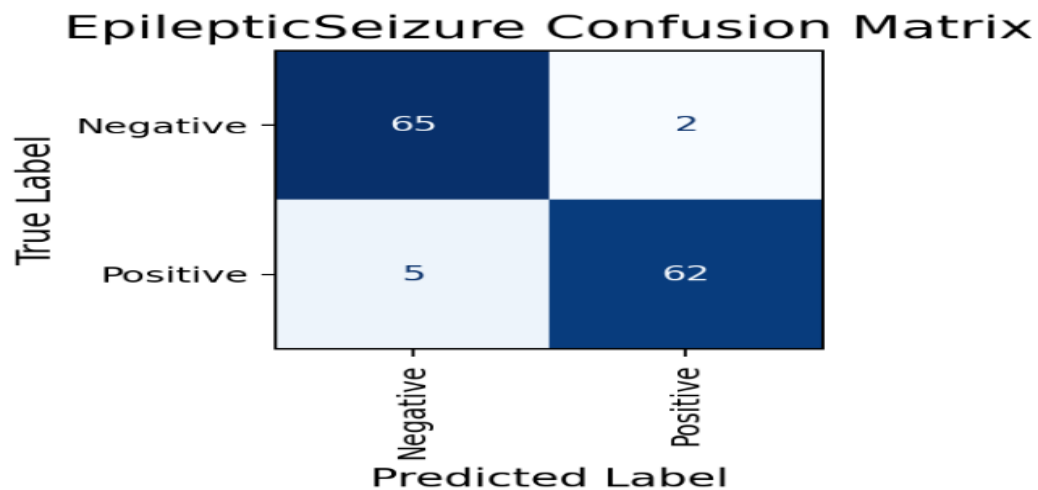


Fig-5: Confusion Matrix for Epileptic Seizure Detection

4.4 Confusion matrix for sleep stage classification:

Figure 6 illustrates the confusion matrix obtained for sleep stage classification. The high concentration of correctly classified samples along the diagonal confirms the ability of the proposed framework to distinguish between different sleep stages with high reliability. Only a small number of samples were incorrectly classified, indicating strong classification performance.

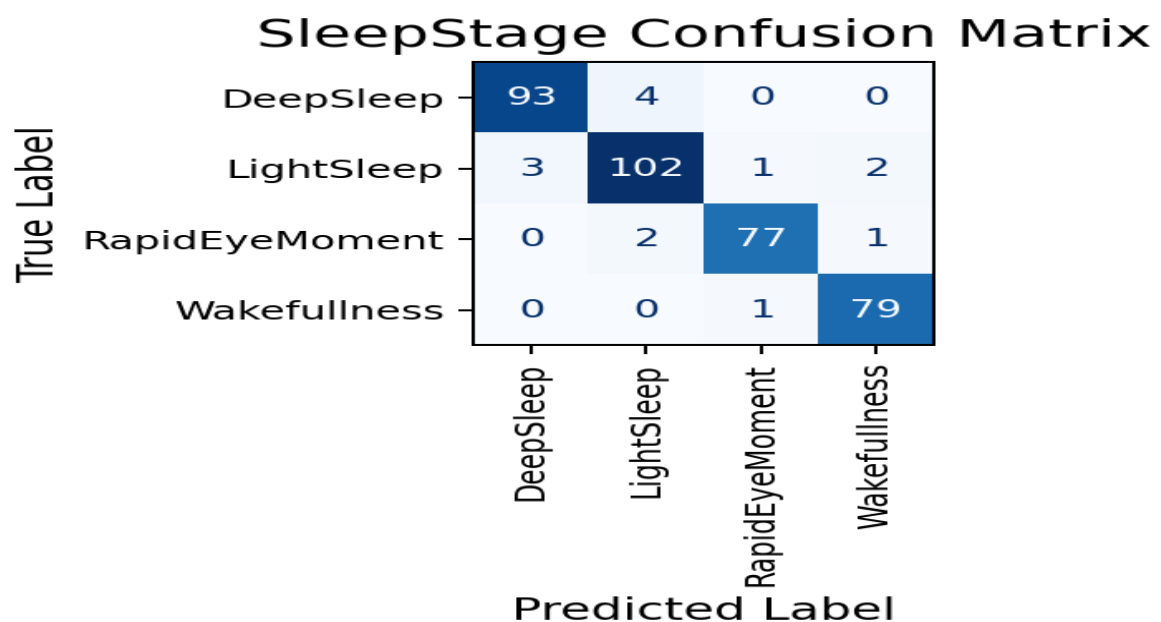


Fig-6: Confusion matrix for sleep stage classification.

4.5 Confusion Matrix for Emotion Recognition:

The confusion matrix shown in Figure 7 demonstrates the effectiveness of EfficientNet-B3 in recognizing facial emotions. Most emotional expressions are correctly identified with minimal overlap between similar facial expressions. The results indicate that the deep feature extraction capability of EfficientNet-B3 effectively captures subtle facial characteristics.

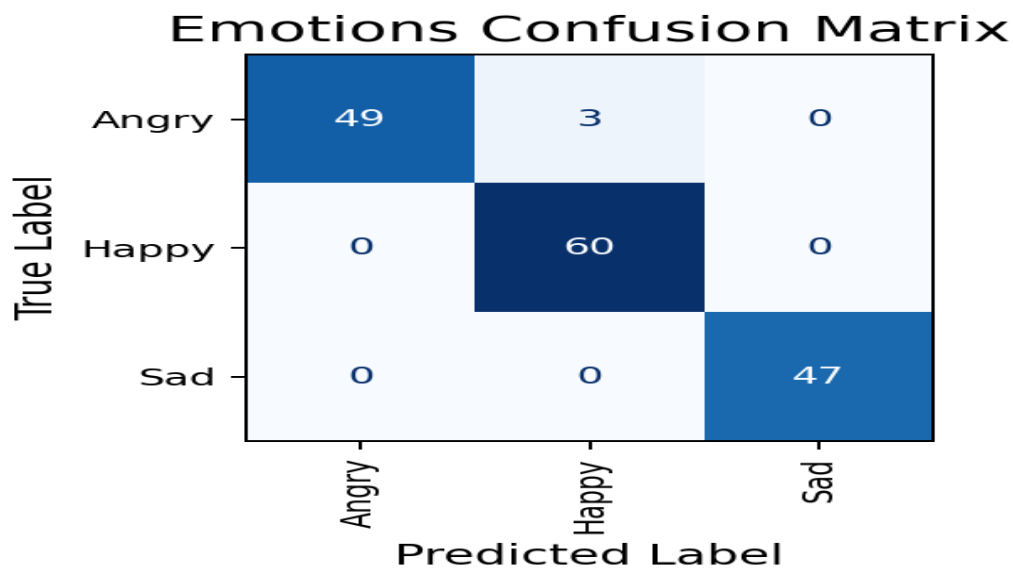


Fig-7: Confusion Matrix for Emotion Recognition

4.6 Confusion Matrix for Skin Cancer Classification:

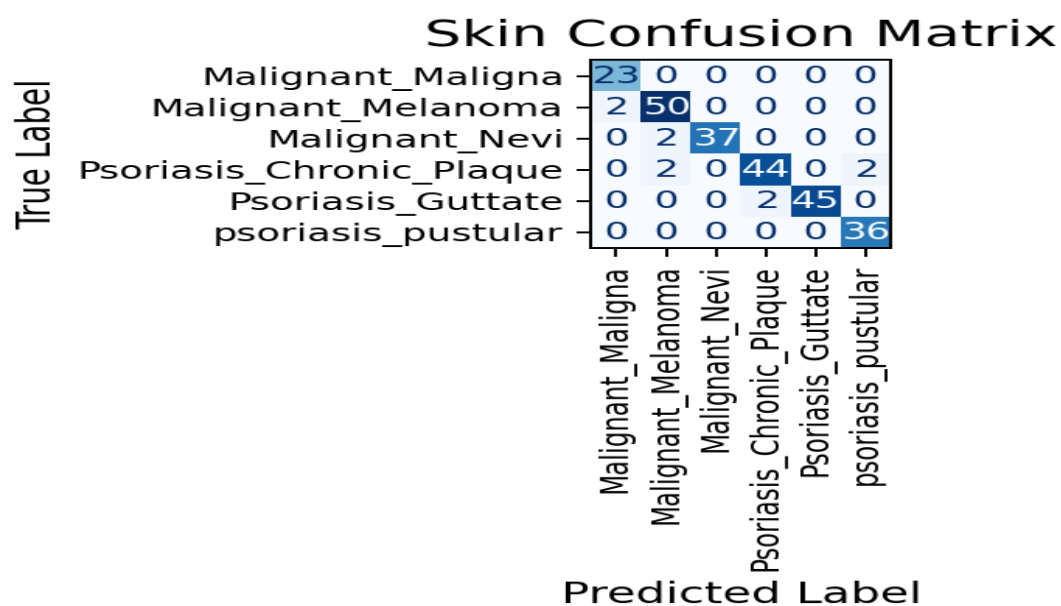


Fig-8: Confusion Matrix for Skin Cancer Classification

Figure 8 presents the confusion matrix for skin cancer classification. The majority of skin lesion images were accurately classified into their corresponding categories. The small number of misclassified samples indicates the robustness of the proposed model in extracting significant dermatological features, resulting in high diagnostic accuracy.

4.7 Performance Comparison Across Different Datasets

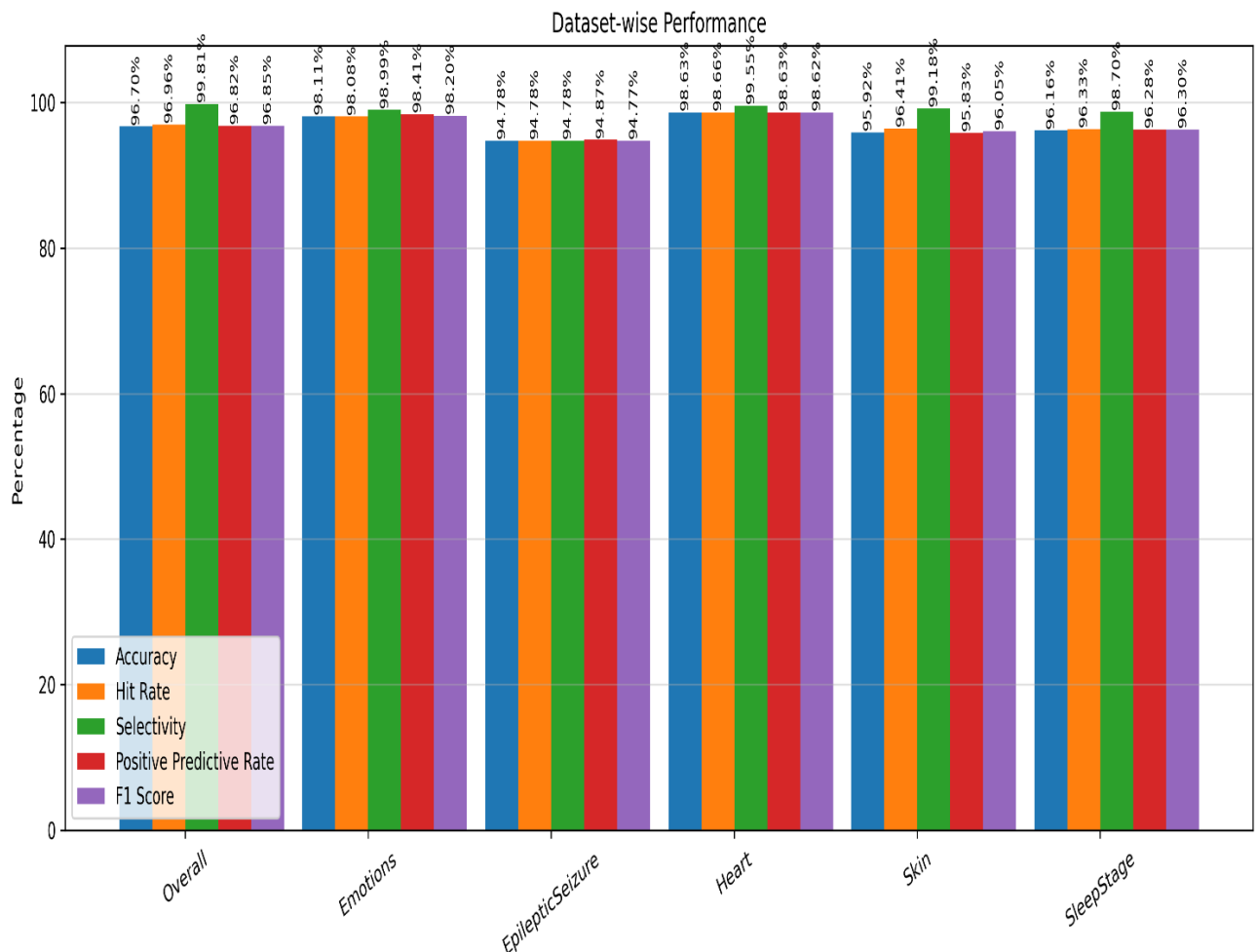


Fig-9: Performance Comparison Across Different Datasets

Figure 9 presents the classification accuracy obtained on different biomedical datasets, including ECG, EEG (epileptic seizure), EEG (sleep stage), skin cancer, and facial emotion datasets. The proposed framework consistently achieved high accuracy across all datasets, demonstrating its capability to generalize effectively to heterogeneous medical image classification tasks. The results indicate that the proposed model provides robust performance regardless of differences in image characteristics and class distributions.

4.8 Performance Comparison of Different CNN Models:

Figure 10 compares the classification performance of three different deep learning approaches, namely DenseNet-SVM, NASNet-XGBoost, and the proposed EfficientNet-B3 model. The proposed EfficientNet-B3 achieved the highest classification accuracy of 96.7%, outperforming NASNet-XGBoost (95%) and DenseNet-SVM (91%). The improved performance is mainly attributed to EfficientNet-B3's compound scaling strategy and transfer learning capability, which enable effective extraction of discriminative features while maintaining computational efficiency.

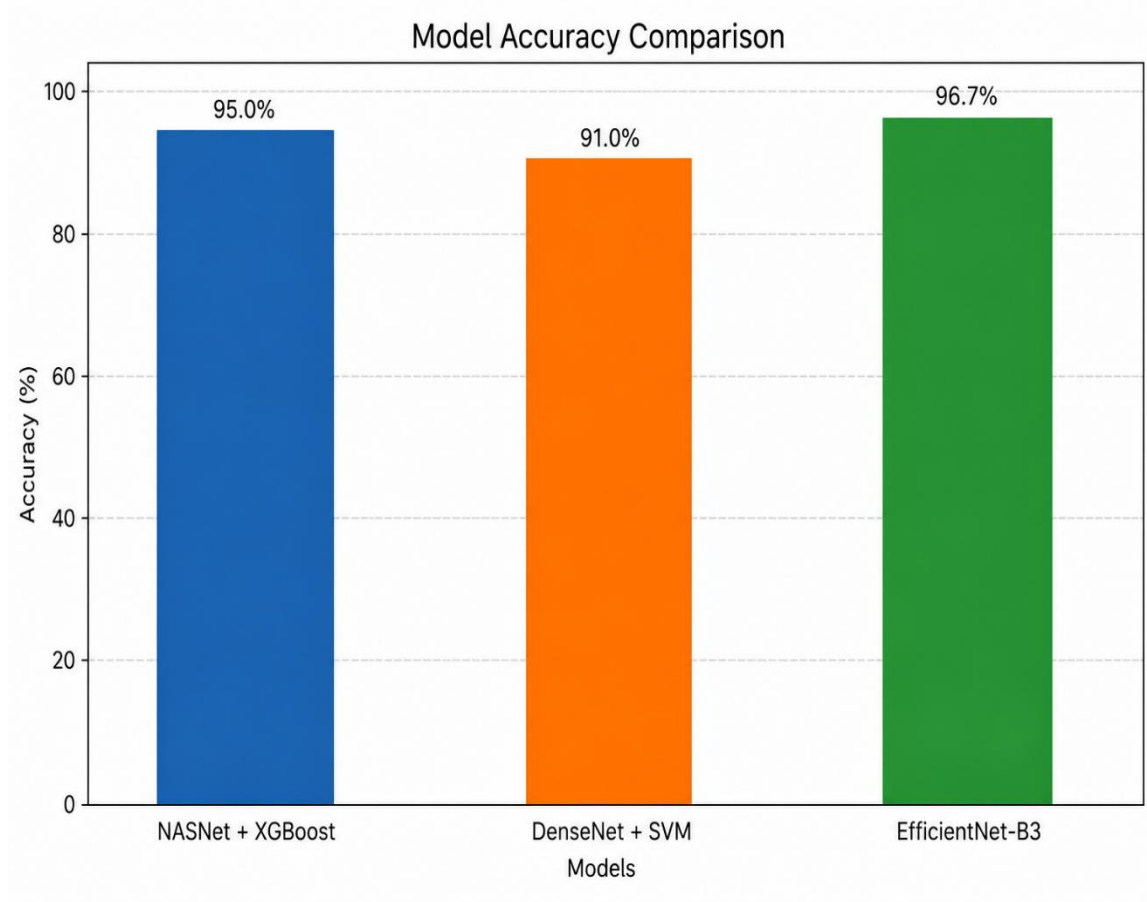


Fig-10: Performance Comparison of Different CNN Models

Table-1 compares the performance of the proposed EfficientNet-B3 model with NASNet-XGBoost and DenseNet-SVM using four evaluation metrics, namely Accuracy, F1-score, Hit Rate, and Selectivity. The proposed EfficientNet-B3 achieved the highest classification accuracy of **96.7%**, outperforming NASNet-XGBoost (**95.4%**) and DenseNet-SVM (**91.0%**). Similarly, the proposed model attained an **F1-score of 96.2%**, indicating a balanced

improvement in both precision and recall. Furthermore, the EfficientNet-B3 model achieved a **Hit Rate of 96.5%** and **Selectivity of 97.1%**, demonstrating its superior capability in correctly identifying both positive and negative samples. These results confirm that EfficientNet-B3 provides more robust feature extraction and better generalization across multiple biomedical datasets than the existing comparison models.

Performance Metric (%)	Proposed EfficientNet-B3	NASNet-XGBoost	DenseNet-SVM
Accuracy	96.7	95.4	91.0
F1-Score	96.2	94.1	89.8
Hit Rate (Sensitivity/Recall)	96.5	94.7	92.3
Selectivity (Specificity)	97.1	96.3	93.8

Table-1: Performance Comparison of Different Classification Models

V. CONCLUSION

The research studies biomedical image classification through its application of EfficientNet-B3 architecture as the standard backbone for the study. The system introduces a dynamic terminal-based key selection mechanism to enable multi-domain operation through a single trained model which eliminates the requirement for different architectures that handle multiple modalities. The unified design approach achieves two advantages by decreasing the need for duplicate computing resources and making it easier to deploy systems in real-world healthcare settings. The experimental results demonstrates that the framework achieves 96.30% accuracy

across ECG, EEG, facial image datasets while using compound scaling and transfer learning for efficient computation. The results demonstrate strong generalization abilities across various biomedical input structures. Future work may explore extending the frame work to additional medical imaging modalities, integrating multimodal data fusion strategies, and optimizing the architecture for real-time embedded or edge-based healthcare systems.

REFERENCES

- [1] S. Ryu and I. Joe, "A hybrid DenseNet-LSTM model for epileptic seizure prediction," *Applied Sciences*, vol.11, no.16, p.7661, Aug.2021.
- [2] S. M. Usman, S. Khalid, and M. H. Aslam, "Epileptic seizures prediction using deep learning techniques," *IEEE Access*, vol.8, pp.39998-40007, 2020.
- [3] A. Arza, J. M. Garzon-Rey, J. Lazaro, E. Gil, R. lopez-Anton, C. de la camara, P. Laguna, R. Bailon, and J. Aguilo, "Measuring acute stress response through physiological signals: Towards a quantitative assessment of stress," *Medical & Biological Engineering & Computing*, vol.57, no.1, pp.271-287, Jan.2019.
- [4] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE Access*, vol.7, pp.111866-118878, 2019.
- [5] P. N. Srinivasu, U. Sirisha, K. Sandeep, S. P. Praveen, L. P. Maguluri, and T. Bikku, "An interpretable approach with explainable AI for heart stroke prediction," *Diagnostics*, vol.14, no.2, p.128, Jan.2024.
- [6] K. A. Hossain, "Analysis of present and future use of artificial intelligence (AI) in line of fourth industrial revolution (4IR)," *Scientists Research Journal*, vol.11, no.8, pp.1-50, Aug.2023.
- [7] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Towards in sighting cybersecurity for healthcare domains: A comprehensive review of recent practices and trends," *Cyber Security Applications*, vol.1, Art.no.100016, Dec.2023.
- [8] S. Ponnada, T. K. Tak, P. R. Kshirsagar, P. S. Rao, and A. Dayal, "Expanding applications of TinyML in versatile assistive devices: From XGBoost transfer learning," *IEEE Access*, 2024, doi:10.1109/Access.2024.3496791