

# Chapter 5

## Result Analysis

### 5.1 Performance Evaluation

For the evaluation of performance of our model, we have used some metrics to determine how accurately our model can classify the images according to the correct label:

#### Accuracy:

It refers to ratio of total correct predictions across both positive and negative classes to the total number of predictions. The formula for accuracy is,

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

Here, TP is True Positive

TN is True Negative

FP is False positive

FN is False Negative

#### Precision:

It measures the correctly predicted positive samples out of all positive predictions made by the model. Formula for calculating precision is,

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

#### Recall:

Recall score refers to the accurate predictions the model made for all positive samples and it is also known as sensitivity score. Recall is calculated as,

$$\text{Recall} = \frac{TP}{TP + TN} \times 100\%$$

#### F1- Score:

It is the harmonic mean of the precision and recall and its formula is,

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$$

#### Out-of-bag (OOB) Score:

OOB score is used for random forest classifiers to measure how accurately it can classify samples it has not been trained on and it can be used as a validation set.

### Confusion matrix:

This matrix is a visual representation of the amount of samples classified in the classes as true predictions and false predictions.

After evaluating the model, it is observed that the model can make accurate predictions with an accuracy rate of 77.92%, which is a very good metric for neuromorphic models trained with unsupervised SNN. Our model has achieved a weighted precision score of 79.47% and macro precision of 74.49%. Weighted precision is more important for our model as there is imbalance in the dataset, considering that the number of demented cases is much higher than nondemented cases. The recall and F1 score are 77.92% and 78.41% respectively.

Table 5.1: Overall Performance Evaluation

Metrics	Overall
Accuracy	0.7792
Precision(weighted)	0.7947
Precision(macro)	0.7449
Recall(weighted)	0.7792
F1-Score(weighted)	0.7841
OOB Score	0.7769

Table 5.2: Overall System Summary

Ensemble Components	Accuracy
Random Forest	0.7877
Gradient Boosting	0.7693
Ensemble	0.7792

We have implemented two classifiers in the model and then used an ensemble classifier for better performance. The random forest classifier performs well with a 78.77% accuracy and OOB score of 77.69%. Gradient boosting achieves an accuracy of 76.93%. Finally, the accuracy of the ensemble classifier for our model stands at 77.92%.

## 5.2 Analysis of Design Solutions

The model can contribute greatly to the medical field as an aid for doctors to get a second opinion about their diagnosis as computers can capture patterns that might not be visible to human eyes. As it can be observed from our model performance evaluation, the model performs quite well in binary classification and can detect demented samples accurately almost 78% of the time. If this model is used by healthcare professionals in hospitals or diagnostic centers, it can reduce the burden on healthcare professionals for manually interpreting each MRI scan.

As the model can be implemented using open source software and standard hardware and very expensive neuromorphic hardware or costly softwares are not required for the implementation, the model can be considered cost effective for the benefits it can provide in the field of healthcare. The model's accuracy might seem lower compared to some traditional deep learning models because it uses SNN for imitation of brain activity and uses mechanisms without any backpropagation. The model learns the features without any backpropagation and extracts features from the images using unsupervised autoencoder structure.

The model is implemented without any backpropagation of data, loss or learning features and can still achieve a high accuracy. This approach requires reduced computations, less training time, it saves memory and is more energy efficient. It does not require very complicated advanced hardware, reducing carbon emissions and therefore, this model can be implemented with reduced carbon footprint. Although the accuracy might seem low compared to traditional deep learning models, the model's imitation of brain neuronal activity by producing spikes is suitable for capturing neuronal activity patterns and structural patterns. So, it is more feasible for classification of diseases relating to the brain, like dementia and Alzheimer's.

Considering the reasonable costs and energy efficiency of the model, the accuracy can be considered a fair trade-off for the implementation of this model. However, the model might have some areas of improvement to enhance the accuracy by extracting even higher quality features. This can be done by using a branch for the autoencoder which will extract high quality features using an established CNN model as CNN models usually perform well on image classification and feature extraction. However, it uses backpropagation to propagate the loss or error signals backward from the output layer. Therefore, the benefits of reduced computation, reduced carbon footprint and energy efficiency is lost with a dual branch autoencoder model implementing CNN.

### 5.3 Final Design Adjustments

From the perspective of performance evaluation, our model has some limitations due to the unsupervised training nature of Spiking Neural Networks. If more emphasis is placed on precision and accuracy, more accurate feature extraction is essential. So we have developed another model, with our SNN autoencoder as the baseline model and implemented a CNN pretrained feature extractor branch for extracting higher quality features and better performance.

In our enhanced dual branch model incorporating CNN which leverages the advantages of backpropagation instead of following the biological neuronal activity and structure, we have used two autoencoder branches for feature extraction, one branch uses our Forward-Forward and R-STDP integrated SNN autoencoder and the other branch uses pretrained ResNet50 backbone for CNN feature extraction.

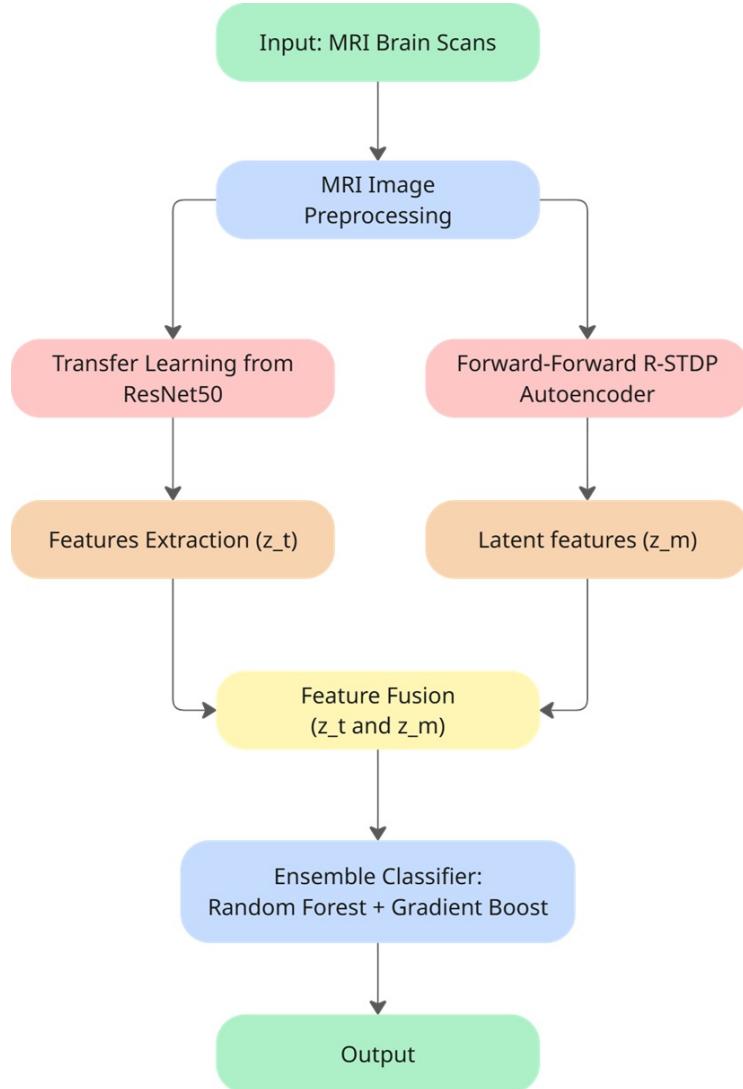


Figure 5.1: Overview of Fusion model

The MRI images of  $224 \times 224$  dimensions are preprocessed in the same way as our base model. Then the MRI images are fed into the ResNet50 architecture, some of the layers are frozen and the other layers and parameters are trainable. Then 512 dimensional features are extracted from the model and using Transfer Learning it is fed into SNN architecture from where the feature vector,  $(z_t)$  is obtained. The input MRI images are also separately fed into our SNN autoencoder structure and it extracts the latent 256 dimensional  $(z_m)$  vector. Then the vectors  $(z_m)$  and  $(z_t)$  are concatenated for fusion. The two models can extract complementary features, and after fusion, the ensemble classifier can train on the fused features, thus improving the classification.

Table 5.3: Performance Evaluation of Dual Branch Fusion Model

<b>Metrics</b>	<b>Overall</b>
Accuracy	0.8709
Precision(weighted)	0.8698
Precision(macro)	0.8509
Recall(weighted)	0.8709
F1-Score(weighted)	0.8702
AUC score	0.9361

The model can be evaluated using an ensemble classifier after training and classification performed on the same test data as our original model. It can be observed that the ensemble accuracy of the model has now increased to 87.09%. This improvement in accuracy is due to the complementary feature extraction of the dual branch feature extractor autoencoder model. The precision, recall and F1 score of the model are 86.98%, 87.09% and 87.02% respectively.

## 5.4 Statistical Analysis

A table is given below showing the performance evaluation of our proposed model:

Table 5.4: Overall and Per-Class Performance Summary

<b>Metrics</b>	<b>Overall</b>	<b>NonDemented</b>	<b>Demented</b>
Accuracy	0.7792	0.738	0.797
Precision(weighted)	0.7947	0.616	0.873
Recall(weighted)	0.7792	0.738	0.797
F1-Score(weighted)	0.7841	0.672	0.834

From table 5.4, it can be observed that our model gave positive results and achieved overall high accuracy of 77.92% using the FF and RSTDp integrated autoencoder. From the table, we can see that the accuracy and recall score remains same and the model has nearly 80% precision, which means it can predict positive samples for demented class fairly accurately.

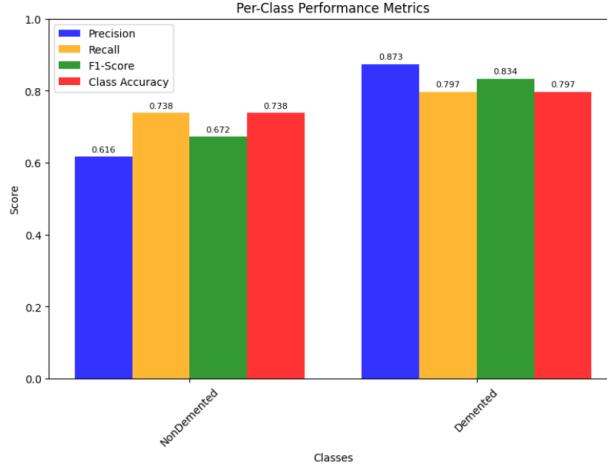


Figure 5.2: Model performance summary

Figure 5.2 shows the precision, recall, f1 score and class accuracy for both the classes showing better results for demented class. The demented class achieved 87.3% precision and 79.7% accuracy in predictions but the non demented class achieved only 61.6% precision and 73.8% accuracy in predictions. This disparity in accurate predictions might be attributed to the imbalance in samples of the two classes. The dataset contains a very high number of demented samples and a much lower number of non demented samples compared to the demented class.

Figure 5.3 shows the confusion matrix generated for raw counts and the normalized confusion matrix generated for our model on the testing data.

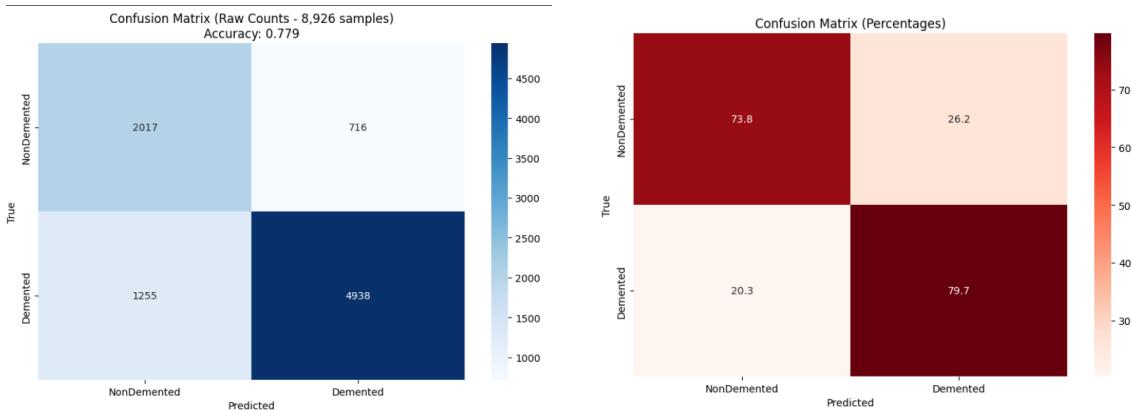


Figure 5.3: Confusion matrix of SNN autoencoder model

Analyzing the matrix, we can observe that the model can accurately predict 2017 non demented samples or 73.8% non demented samples correctly and inaccurately predict 716 non demented samples as demented or make wrong predictions for 26.2% non demented samples. Moreover, the model correctly predicts 4938 demented samples or 79.7% demented samples correctly but inaccurately labels 1255 demented samples as non demented or makes wrong class predictions for 20.3% demented samples.

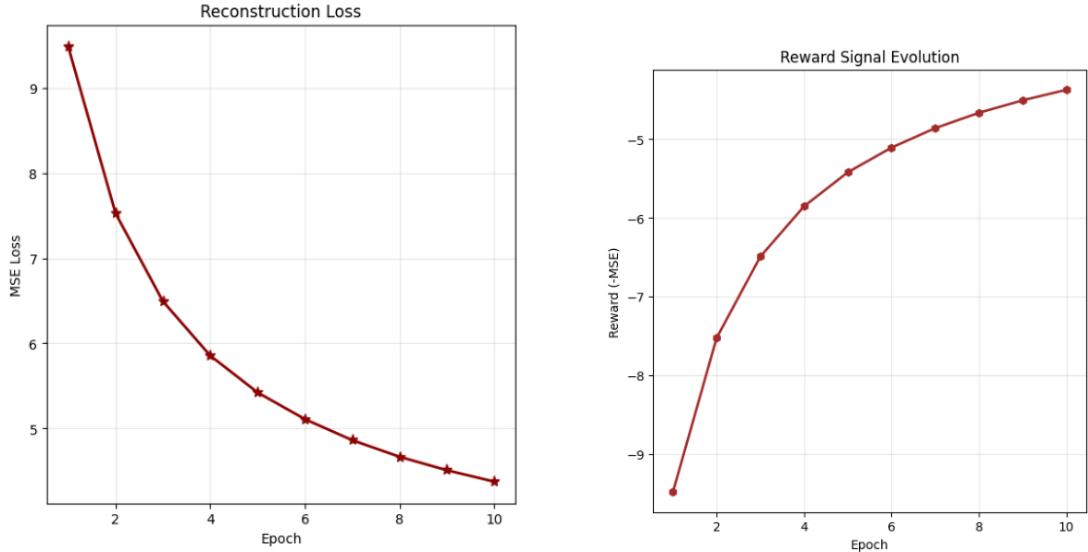


Figure 5.4: MSE loss and rewards per epoch

From the epoch by epoch training graphs for the model in figure 5.4, it can be seen that the reconstruction loss is gradually decreasing over time and the reward or negative MSE is increasing over time. So, the reconstruction loss and reward is inversely proportional. The better the quality of reconstruction, the higher the reward is.

Moreover, from the performance evaluation of the improved model for feature extraction, it can be observed that the accuracy has increased by nearly 10% as the accuracy has jumped from around 77% to around 87%. From the performance summary of both models, a pattern can be observed that the overall accuracy and recall score for the models are equal.

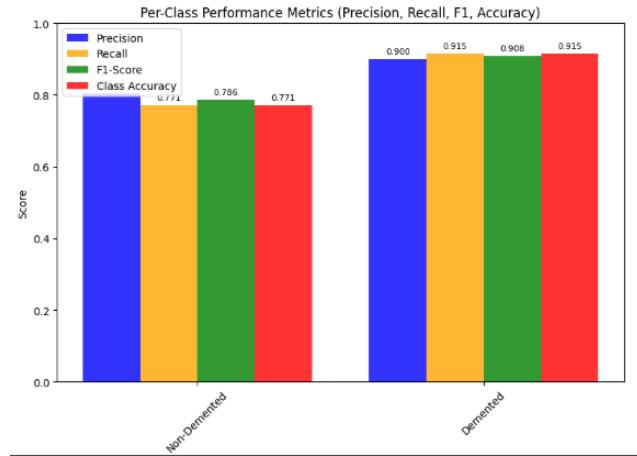


Figure 5.5: Fusion model performance summary

Figure 5.5 shows the precision, recall, F1 score and accuracy for both the classes showing better results for demented class even for the improved model. This class disparity remains despite better feature extraction because of the class imbalance in the dataset. The demented class achieved 90% precision and 91.5% accuracy but

the non demented class has 80.2% precision and 77.1% accuracy for predictions.

Figure 5.6 shows the confusion matrices generated for our fusion model during evaluation.

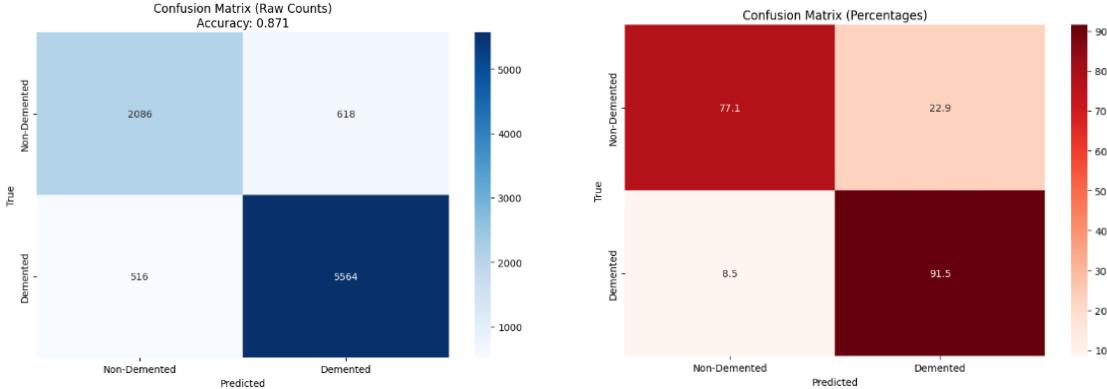


Figure 5.6: Confusion matrix of dual branch fusion model

We can observe that the model has successfully predicted 2086 non demented samples with 77.1% accuracy and incorrectly predicted 618 non demented samples as demented or made wrong predictions for 22.9% non demented samples. The model has also correctly predicted 5564 demented samples with 91.5% accuracy but inaccurately labelled 516 demented samples as non demented or inaccurately classified 8.5% demented samples.

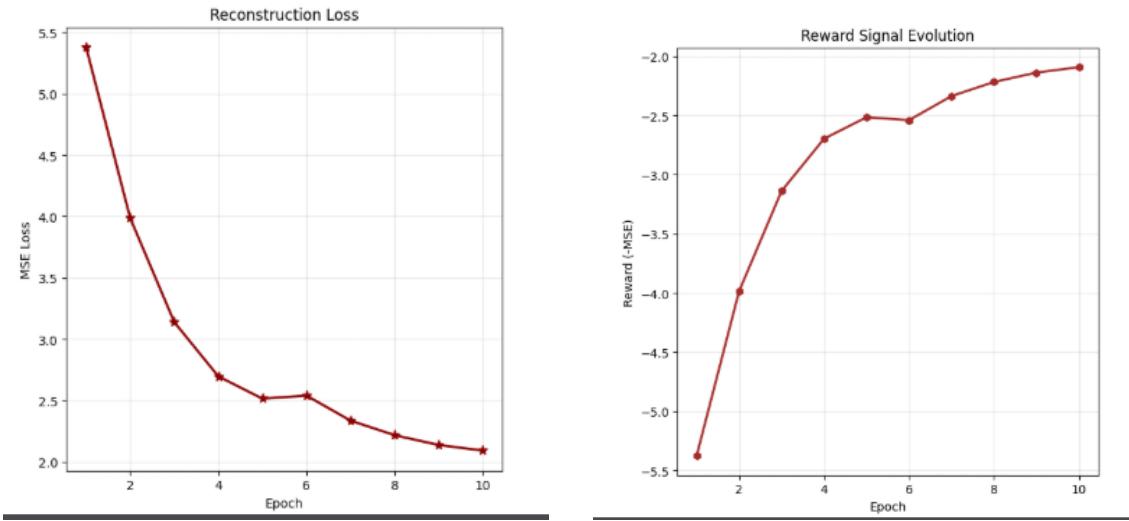


Figure 5.7: MSE loss and rewards per epoch

From the loss vs epoch and reward vs epoch training graphs in figure 5.7, it can be seen that the reconstruction loss is gradually decreasing over time and the reward is increasing over time but the loss slightly increases at epoch 6 and the reward decreases. However, the loss decreases and the reward increases again after epoch 6, demonstrating an inverse relationship between the two parameters.

## 5.5 Comparisons and Relationships

Based on the coding approaches across five different models for Alzheimer’s disease classification, the table given below presents a comprehensive performance comparison:

Table 5.5: Comparative Performance Analysis

Model	Precision	Recall	F1 Score
ViT-B/16 (Vision Transformer)	0.9928	0.9928	0.9928
ResNet-50 (CNN)	0.9783	0.9821	0.9788
SNN with Surrogate Gradient	0.8677	0.8703	0.8678
Proposed Model Fusion with ResNet-50	0.8698	0.8709	0.8702
FF R-STDP Autoencoder (Proposed)	0.7947	0.7792	0.7841
SNN with STDP & FF	0.3475	0.3646	0.3475
Unsupervised STDP SNN	0.0196	0.1398	0.0343

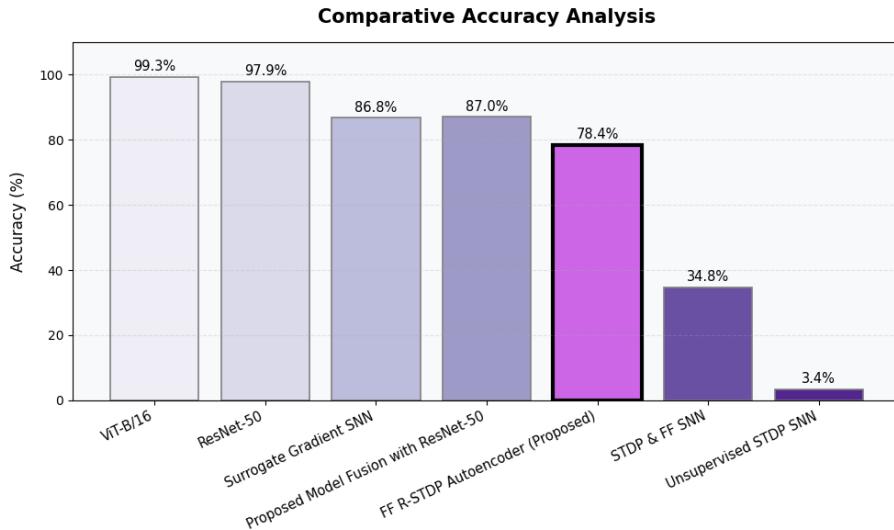


Figure 5.8: Bar chart illustrating the comparative accuracy of different models

The results demonstrate a clear inverse relationship between biological plausibility and classification accuracy. ViT-B/16 that is pre-trained on ImageNet-21k achieves the highest accuracy 99.3%, but requires backpropagation and has minimal energy efficiency than SNN models. In contrast, the unsupervised STDP approach, while most biologically plausible, achieves only 3.4% accuracy, highlighting the limitations of purely unsupervised learning for medical image classification.

Our proposed model, The FF R-STDP Autoencoder provides feedback that enables the network to learn discriminative features, improving accuracy over Forward-Forward and pure STDP. However, while applying hyperparameter tuning, especially increasing epochs during training, it tended to exhibit overfitting behavior. The dual-branch fusion model combines the strengths of neuromorphic feature extraction and CNN-based representations. This improvement over the pure neuromorphic

approach demonstrates that hybrid architectures can balance biological plausibility with practical performance requirements.

## 5.6 Discussions

This research highlights the novelty that biologically-inspired neuromorphic computing can serve as a viable alternative to traditional Deep Learning (DL) for medical image classification, though with important tradeoffs between biological plausibility, computational efficiency, and classification accuracy. The performance hierarchy observed across the models reveals a fundamental tension in neuromorphic computing. Pure unsupervised STDP represents the most biologically faithful approach but proves inadequate for clinical applications. The brain itself uses reward-driven learning through dopaminergic modulation, making R-STDP arguably more biologically accurate than pure STDP despite being more complex. Unlike conventional CNN features that capture spatial patterns, these temporal features encode intensity information as firing dynamics.

Our final model consistently performed better on the Demented class compared to the Non-Demented class. This disparity, while partially attributable to dataset imbalance, raises important questions about clinical utility. In screening scenarios, brain atrophy patterns (spatial) and tissue intensity variations (encoded temporally through spike timing) both contribute to diagnostic accuracy and detecting false negatives (missing actual dementia cases) may be more costly than false positives. Our models' higher recall for the demented class suggests they would function reasonably as screening tools, though the lower precision for non-demented cases indicates a risk of overdiagnosis. The weighted sampling approach employed during training mitigates but does not eliminate this imbalance.

The limitations in our model occur because firstly, the model was trained and tested primarily on a secondary dataset with binary classification (demented vs. non-demented), while clinical practice requires distinguishing between multiple dementia subtypes (Alzheimer's, vascular dementia, frontotemporal dementia, etc.). Second, while primary hospital data was collected, the quantity remains limited compared to the secondary dataset, raising questions about generalization to patient populations. Moreover, unlike the MNIST dataset, extensive image preprocessing was not performed due to limited computational resources.

# Chapter 6

## Conclusion

### 6.1 Summary of Findings

This research presented a novel neuromorphic approach to classifying Alzheimer’s disease and dementia subtypes using MRI brain scans. The methodology integrated biologically inspired temporal spike encoding with Forward-Forward learning and Reward-Modulated Spike-Timing-Dependent Plasticity (R-STDP) in a three-stage pipeline. The baseline Forward-Forward R-STDP autoencoder model achieved 77.92% ensemble accuracy on the binary classification task (Demented vs. NonDemented), with 79.14% weighted precision and 77.08% out-of-bag score. The neuromorphic preprocessor successfully generated 8,448-dimensional feature vectors capturing temporal dynamics from  $32 \times 32$  MRI images, which were compressed to 256 dimensions through the autoencoder’s latent bottleneck. Advanced feature selection techniques reduced the combined feature space to 2,179 optimal dimensions.

An enhanced dual-branch architecture incorporating both the SNN autoencoder and a pretrained ResNet50 CNN feature extractor demonstrated significant performance improvement, achieving 87.09% ensemble accuracy with 86.98% precision and 87.02% F1-score. This improvement validated the complementary nature of the neuromorphic temporal features and the traditional spatial features of CNN. Both models showed higher accuracy for the demented class (80.5% and 91.5% respectively) compared to the non-demented class (71.8% and 77.1% respectively), attributed to class imbalance in the dataset. The reconstruction loss consistently decreased across training epochs while reward signals increased, confirming the effectiveness of the R-STDP mechanism in guiding feature learning without backpropagation.

### 6.2 Contributions to the Field

In our research, we designed a novel spiking autoencoder for the classification of dementia from structural MRI scans. This research makes significant contributions to the intersection of neuromorphic computing and medical image analysis. In particular, we propose an energy efficient and biologically plausible solution for the detection of dementia from only structural MRI scans with a high accuracy. The contributions our research have made to the field is summarized as follows:

- We designed an energy efficient approach for the classification of dementia

and early diagnosis of Alzheimer’s, which can be used as an aid for medical healthcare professionals in diagnosis.

- The framework provides healthcare professionals with a practical tool that can serve as a decision support system for dementia detection, potentially reducing the burden of manual MRI interpretation and providing a second opinion that captures patterns, which are not immediately visible to human observers.
- The Forward-Forward algorithm optimizes each layer independently, while R-STDP modulates synaptic weights according to spike timing and reconstruction quality. This combination provides a biologically plausible alternative to conventional DL that could inspire future neuromorphic architectures.
- The development of a comprehensive neuromorphic feature extraction pipeline represents another key contribution. The temporal encoding scheme extracts distinct spike-based features that capture both statistical properties and temporal dynamics of neuronal firing patterns.
- The validation of a hybrid neuromorphic-CNN architecture demonstrates that these seemingly disparate approaches can be synergistically combined.
- The dual-branch model achieving higher accuracy proved that neuromorphic temporal features and traditional CNN spatial features provide complementary information that enhances overall classification performance.
- We designed the model and trained it on a secondary dataset which was not collected from local people but tested it on a primary dataset collected from a hospital in Bangladesh. Simulations were performed on many different subsets, entire datasets and using different hyperparameters and our model showed good performance. This validates the robustness of our model and ability to perform well on different datasets.

### 6.3 Recommendations for Future Work

The model can be expanded to multi-class classification distinguishing between Mild Cognitive Impairment, varying dementia severities, and other neurodegenerative conditions such as Parkinson’s and Frontotemporal dementia. A particularly promising avenue involves multi-modal integration using synchronous MRI-EEG acquisitions. Simultaneous scanning captures both visible structural brain changes in MRI and functional neural activity patterns. This is because Early-stage dementia often shows EEG abnormalities before substantial structural changes are visible, so combining modalities could enable earlier detection than considering either of them alone. The model could be enhanced further in future for even better feature extraction without incorporating CNN if even better algorithms for SNN without backpropagation are discovered as it is still an emerging field of neural networks. An enhanced model with a better classification accuracy could aid medical professionals greatly to diagnose neurodegenerative conditions with an energy efficient approach.