International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Improved Accuracy for ECG Arrhythmia Classification using AlexNet-M-SE with Dropout and BatchNorm

Krishna Chaitanya Pidugu[1], Dr. Manjula Devarakonda Venkata[2]

[1,]*Department of CSE, Pragati Engineering College(A)*
[2]*Department of CSE, Pragati Engineering College(A)*

---***---

Abstract - AlexNet—released by Krizhevsky, Sutskever, and Hinton in 2012—initiated the deep-learning revolution with its win in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2012) by an unprecedented error reduction over traditional methods. The five-convolutional-layer (informally referred to as "AlexNet-5 conv") model applies max-pooling, ReLU non-linearities, local response normalization (LRN), data augmentation, dropout, and stochastic gradient descent (SGD) with momentum—every one of them trained on GPUs. This paper outlines the architecture, training pipeline, and contributions; discusses shortcomings; This paper presents future directions and open problems. Overall, AlexNet-5 is best understood as an early template whose ideas—ReLU, aggressive augmentation, dropout, and GPU-first training—remain at the core of current architectures even as the field continues to move toward deeper, normalized, and efficiency-optimized networks.

Key Words: Convolutional Neural Networks, AlexNet, ReLU, Dropout, ImageNet, GPU Training, Deep Learning.

1.INTRODUCTION

Prior to 2012, vision recognition was based on hand-crafted features (SIFT/HOG) with shallow classifiers. AlexNet demonstrated that end-to-end learned features could handily trounce them on large, labeled sets like ImageNet (1.2M images, 1,000 classes) [1],[2]. AlexNet has five convolutional layers and three fully connected layers, so the standard shorthand "AlexNet-5 conv." The original implementation also divided the model across two GPUs to conserve memory and speed up training [1].

Key Contributions:

• ReLU activations enabling fast, non-saturating training [1],[3].
• Max-pooling overlap improving translational robustness.
• Local response normalization (LRN) as a then-useful regularizer [1].
• Dropout in fully connected layers to prevent overfitting [4].
• Large data augmentation (random crop, flip, color jitter) [1].
• GPU training for end-to-end multi-epoch optimization at scale.

Table-I organizes AlexNet's forward pass layer by layer for an input 224×224 RGB image—kernel/stride/pad, resulting feature-map sizes step by step, and the primary operations (ReLU, LRN, max-pooling, dropout). It illustrates how spatial resolution decreases and channel depth increases from Conv1–Conv5, followed by FC6–FC8 and Softmax—a simple plan to execute and replicate.

Stage	Layer (kernel/stride/pad)	Maps × size	Notes
Input	—	3×224×224 (originally 227×227)	RGB image
Conv1	11×11 / 4 / 2	96×55×55	ReLU → LRN → 3×3 max-pool / 2
Conv2	5×5 / 1 / 2	256×27×27	ReLU → LRN → 3×3 max-pool / 2
Conv3	3×3 / 1 / 1	384×13×13	ReLU
Conv4	3×3 / 1 / 1	384×13×13	ReLU
Conv5	3×3 / 1 / 1	256×13×13	ReLU → 3×3 max-pool / 2
FC6	—	4096	ReLU + Dropout (p=0.5)
FC7	—	4096	ReLU + Dropout (p=0.5)
FC8	—	1000	Softmax

Table-1: AlexNet (5-Conv) Architecture Summary

The network's backbone consists of five large early convolutional layers with large early receptive fields (e.g., 11×11 stride-4 in Conv1) followed by later 3×3 smaller kernels. The three main ingredients are rectified linear unit (ReLU) activations, overlapping max-pooling, and local response normalization (LRN) as initially introduced. Three fully connected layers (two of 4096 units) finish off the classifier, with dropout (p=0.5) to prevent overfitting. The training recipe used large-scale data augmentation—random crops, horizontal left-right reflection, and color perturbations—plus stochastic gradient descent with momentum, weight decay, and stepwise learning-rate schedules. To keep the workload tractable in 2012, the model and minibatches were divided between two consumer GPUs. Overall, AlexNet used about 60 million parameters and trained on ~1.2 million images resized to 224–227 pixels.

From the deployment perspective, AlexNet's size and compute requirements restricted real-time and edge usage without pruning, quantization, or architect rethinking. The gap analysis in this report compares AlexNet's strengths with contemporary solutions: global average pooling for substituting large fully connected layers, residual connections for facilitating depth, normalization layers for regularizing optimization, and efficiency techniques (depthwise separable kernels, group/pointwise convolutions) for mobile use cases.

2. LITERATURE SURVEY

2.1 Overview of Implications

AlexNet's top-5 error (~15.3%) reduced the then-best by more than 10 percentage points, eliciting immediate follow-on work on deeper, more regularized CNNs (VGG, GoogLeNet, ResNet) and the wholesale use of GPUs [1],[5]–[7]. AlexNet (popularly known as AlexNet-5 for five convolutional layers) was the milestone from hand-engineered features to end-to-end deep learning in vision. Labeled on the 1.2-million-image ImageNet dataset, it brought together large early receptive fields (11×11 stride-4 in Conv1) and small 3×3 kernels later in the network, and combined three fully connected layers to form the classifier. The key engineering choices were ReLU activations for efficient, non-saturating optimization; overlapping max-pooling for translation robustness; local response normalization (LRN) as a then-effective regularizer; aggressive data augmentation (random crops, flipping, color jitter); and dropout (p=0.5) in the fully connected layers to prevent overfitting. Stochastic gradient descent with momentum, weight decay, and stepwise learning-rate schedules were a reliable training recipe. Since 2012 GPUs were memory-constrained, the model and minibatches were split across two devices.

Empirically, AlexNet reduced the ILSVRC-2012 top-5 error by more than ten percentage points, demonstrating that learned hierarchical features generalize much better than SIFT/HOG pipelines. It triggered transfer learning in computer vision: AlexNet pre-trained weights became default detection, segmentation, and fine-grained task initializations. Its success also became the benchmark for GPU-first training and rendered practical approaches—ReLU, strong augmentation, and dropout—default options in many situations. Even though subsequent models (VGG, Inception, ResNet, EfficientNet) outperformed it in accuracy and efficiency, AlexNet set the template and lexicon for contemporary CNN design.

2.2 Practical Challenges Identified

• Model & compute size: ~60M parameters; large FC layers consume most memory and FLOPs.
• Stability in training: Weight-decay sensitive and learning-rate schedules; LRN now outdated in comparison to BN [8].
• Generalization gaps: Large, labeled datasets needed; data bias and domain shift are still issues [2].
•Deployment: The original model is costly for edge/mobile deployment without compression.
• Parameter and compute footprint: At about 60M parameters, the majority of which are from the fully connected head, AlexNet is memory-intensive and not edge or mobile deployable.
• Old normalization: LRN achieved minor improvements but was surpassed subsequently by batch normalization and its variations; in the absence of BN, optimization remains initialization- and learning-rate schedule-sensitive.
• Risk of overfitting in FC layers: Dropout alleviated it, but the deep FC stack still overfits on small datasets and increases inference cost; recent architectures instead employ global average pooling and smaller heads.
• Stability training and tuning load: Performance depends on appropriately selected learning rates, momentum, and decay; absence of BN lowers tolerance to large batches or rapid schedules.
• Explainability: Although saliency techniques can project AlexNet's features, its choices are difficult to reason out for high-stakes use in the absence of domain-specific attribution and auditing.
• Net takeaway: AlexNet concepts initiated the deep vision age, but its size, normalization approach, and shallow depth constrain present performance and deploy ability—problems subsequent architectures explicitly solve.
• Batch Normalization and Dropout [15],[16]: Minibatches for stability. Randomly zeros a fraction to reduce overfitting.
• SE (Squeeze and Excitation) [17] : Blocks global spatial stats and reweight feature maps.

Aspect	AlexNet Strength	Modern Resolution
Nonlinearity	ReLU sped training	BN/LayerNorm [8]
Pooling	Overlapping max-pool	Strided convs, dilations; attention
Regularization	Dropout was effective in FC layers	Dropout mainly in classifier; BN + strong augmentation
Capacity	High capacity for ImageNet	Global average pooling; residuals [7]
Depth	5 conv blocks	Very deep nets with skip connections (ResNet-50/101) [7]
Compute	Two-GPU split- 2012	EfficientNets/MobileNets; pruning/quantization [9]

Table-2: Gap analysis of various aspects

2.3 ECG

Clinically, ECG (ElectroCardioGram) [10] is crucial in the detection of arrhythmias (e.g., atrial fibrillation, AV block, ventricular tachycardia), myocardial ischemia and infarction (through ST-segment and Q-wave changes), disturbances in conduction (bundle branch blocks), chamber enlargement, electrolyte imbalance (e.g., hyperkalemia), and drug effects (e.g., QT prolongation). They direct near-instant decisions—thrombolysis or PCI in ST-elevation MI—pacing and antiarrhythmic therapy monitoring, and perioperative and critical-care follow-up. Serial ECGs assist in monitoring dynamic changes, and correlation with symptoms.

The arrhythmias are classified into 10 classes as below:
#class0 Normal sinus rhythm — Regular rhythm 60–100 bpm of the sinoatrial node with upright P waves in I, II, aVF (usually biphasic in V1), each followed by a narrow QRS, constant PR interval (120–200 ms), and normal QTc. Axis, intervals, and R-wave progression are normal for age. This rhythm is the "reference normal," helpful for comparison with previous ECGs and for ruling out arrhythmias, conduction blocks, ischemia, and chamber overload.

#class1 Sinus tachycardia — Normal sinus rhythm >100 bpm in adults with normal P-axis and 1:1 AV relationship; PR and QRS typically normal. Physiologic causes are exercise, fever, anemia, hypovolemia, pain, anxiety, hyperthyroidism, or medications (e.g., stimulants). Context-dependent meaning:

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

generally a compensatory response; sudden-onset/offset or inappropriateness for environment requires search for SVT or other cause.

#.class2 Sinus bradycardia — Sinus rhythm below 60 bpm with upright P, regular PR in II; found in athletes, sleep, or beta-blockers. Pathologic etiologies are hypothyroidism, ischemia, intracranial hypertension, or sick sinus syndrome. Asymptomatic patient is usually benign; red flags are dizziness, syncope, hypotension, or pauses that suggest sinus node dysfunction or high-grade AV block.

#class3 Left bundle branch block (possible) — Suggestive features: wide QRS (usually ≥120 ms for complete; 110–119 ms for incomplete), absent septal q in V5–V6, I, aVL, wide notched/slurred R in V5–V6; deep rS or QS in V1–V3 with secondary ST-T discordance. LBBB can hide ischemia and signifies intraventricular conduction delay; "possible" denotes borderline morphology or rate-dependent LBBB—take symptoms and echocardiography into account.

#class4 Right bundle branch block (possible) — QRS prolongation with rsR′/rSR′ in V1–V2 ("rabbit ears"), wide terminal S in I and V6, and secondary ST-T changes in right precordials. Incomplete RBBB has QRS 100–119 ms. Usually benign, but can be present with right heart strain (PE), congenital heart disease, or myocarditis; "possible" indicates partial/atypical criteria—see clinical context and comparison ECGs.

#class5 Left ventricular hypertrophy (possible) — Elevated LV voltage (e.g., Sokolow–Lyon: SV1 + RV5/V6 ≥35 mm; Cornell: RaVL + SV3 >28 mm men, >20 mm women) with potential "strain" pattern: lateral ST depression and asymmetric T inversion. Voltage in itself can be normal in young/thin persons; "possible" is applied to borderline criteria. Clinically suggests hypertension, aortic stenosis, or athletic remodeling—corroborate with echo and BP check.

#class6 Right ventricular hypertrophy (possible) — Right axis deviation, tall R in V1 (R/S >1) with progressive diminution across precordium, possible right precordial T-wave inversion ("strain"), and qR in V1. Typically subtle on adults' ECGs; "possible" indicates incomplete criteria. Suspect pulmonary hypertension, chronic lung disease, congenital lesion, or pulmonary embolism; echocardiography assists in establishing RV pressure/size overload.

#class7 ST-elevation MI (possible) — Regional ST elevation in adjacent leads (typical thresholds: in V2–V3 ≥2 mm men ≥40, ≥2.5 mm men <40, ≥1.5 mm women; ≥1 mm in other leads) with reciprocal ST depression, evolving hyperacute T waves and subsequently Q waves. "Possible" implies shape/size not obviously diagnostic or confounders (LBBB, LVH, pericarditis). Treat as time-critical—correlate with onset of pain, troponin, and consider immediate reperfusion if met.

#class8 ST-depression / ischemia (possible) — Horizontal/downsloping ST depression ≥0.5–1 mm in adjacent leads with/without T-wave inversion suggests subendocardial ischemia, demand ischemia, or NSTEMI; may also suggest reciprocal changes. "Possible" indicates borderline or rate-related changes, drug or electrolyte effects (e.g., digoxin). Interpret with symptoms, risk factors, and serial ECG/troponin trends; stress testing or imaging may be indicated. #class9 Atrial/ventricular ectopy (irregular RR, possible) — Irregular RR beats caused by premature beats: PACs have premature and widened P with non-compensatory pause; PVCs are wide, bizarre QRS with no preceding P with often complete compensatory pause; rhythm may be bigeminy/trigeminy or couplets. Benign in the majority but may be a sign of electrolyte disturbance, stimulants, structural disease, or ischemia; high PVC burden or symptoms require ambulatory monitoring and echo.

3. IMPLEMENTATION

3.1 Datasets, Classes

To implement the ECG analysis [11], the dataset[12],[13],[14] in raw format has to be undergone the below process:

- 12-lead dataset in hea, mat, csv formats
- generate a image file from csv file
- classify the images into classes
- classes are as follows:
  - #class0 Normal sinus rhythm
  - #class1 Sinus tachycardia
  - #class2 Sinus bradycardia
  - #class3 Left bundle branch block (possible)
  - #class4 Right bundle branch block (possible)
  - #class5 Left ventricular hypertrophy (possible)
  - #class6 Right ventricular hypertrophy (possible)
  - #class7 ST-elevation MI (possible)
  - #class8 ST-depression / ischemia (possible)
  - #class9 Atrial/ventricular ectopy (irregular RR, possible)
- dataset was executed in three scalable levels
- the execution was distributed as 70/15/15 i.e., 70% was for training, 15% for validation testing and 15% for testing the model
- epoch=20, batch_size=16

Results:

- the dataset was used to test the original AlexNet and modified AlexNet-M model also.

Dataset=204; test_set= 64:

AlexNet Model:

Classification report:

	precision	recall	f1-score	support
Normal sinus rhythm	0.250	1.000	0.400	6
Sinus tachycardia	1.000	1.000	1.000	7
Sinus bradycardia	1.000	1.000	1.000	5
Left bundle branch block (possible)	0.875	1.000	0.933	7
Right bundle branch block (possible)	1.000	1.000	1.000	7
Left ventricular hypertrophy (possible)	0.286	0.286	0.286	7
Right ventricular hypertrophy (possible)	0.000	0.000	0.000	7
ST-elevation MI (possible)	0.000	0.000	0.000	6
ST-depression / ischemia (possible)	0.000	0.000	0.000	6
Atrial/ventricular ectopy (possible)	1.000	1.000	1.000	6
accuracy			0.625	64
macro avg	0.541	0.629	0.562	64
weighted avg	0.541	0.625	0.561	64

Fig-1: Classification report- AlexNet(base)model

AlexNet-M model:

[Test] loss=0.2029 | acc=0.9688

Classification report:

	precision	recall	f1-score	support
Normal sinus rhythm	1.000	1.000	1.000	6
Sinus tachycardia	1.000	1.000	1.000	7
Sinus bradycardia	1.000	0.800	0.889	5
Left bundle branch block (possible)	1.000	1.000	1.000	7
Right bundle branch block (possible)	0.875	1.000	0.933	7
Left ventricular hypertrophy (possible)	1.000	0.857	0.923	7
Right ventricular hypertrophy (possible)	1.000	1.000	1.000	7
ST-elevation MI (possible)	0.857	1.000	0.923	6
ST-depression / ischemia (possible)	1.000	1.000	1.000	6
Atrial/ventricular ectopy (possible)	1.000	1.000	1.000	6
accuracy			0.969	64
macro avg	0.973	0.966	0.967	64
weighted avg	0.973	0.969	0.968	64

Fig-2: Classification report- AlexNet-M-SE model

The above implementation shows that the modified AlexNet-M showed an accuracy of 96.9% while the base model showed only 62.5% accuracy. The Confusion matrix indicates that the AlexNet-M has better recognition capabilities compared to base model.

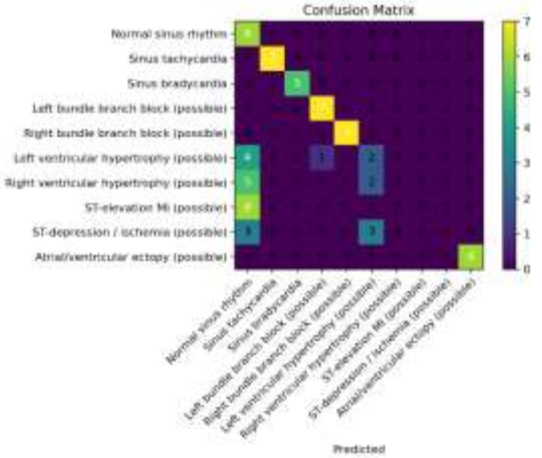


Fig-3: ConfusionMatrix- AlexNet(base)model

AlexNet5 (base): Classes 0–4 and 9 are ideal (all on the diagonal). But LVH (class5), RVH (class6), STEMI (class7), and ST-depression (class8) have systematic false negatives to "Normal"—e.g., most LVH/RVH/STEMI rows concentrate in the first column. Rough estimates from the matrix: class5 ≈14% (1/7), class6 ≈29% (2/7), class7 ≈0% (0/6), class8 ≈50% (3/6). That pattern makes the base model under-call pathology and fall back to NSR when unsure.

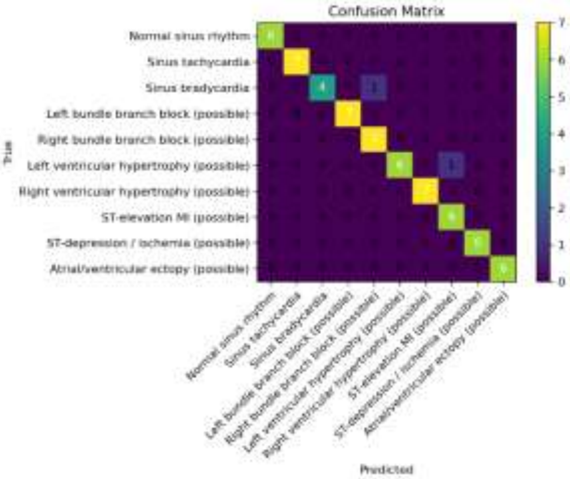


Fig-4: ConfusionMatrix- AlexNet-M-SE model

AlexNet5-M (modified): Nearly all mass is on the diagonal across all classes. Only small residual confusions remain (notably brady vs LBBB and LVH vs RVH), but every class is detected well (e.g., LVH ≈86%, brady ≈80%, others ≈100%). In short: the modified model fixes the "everything looks normal" failure and balances performance class-wise.

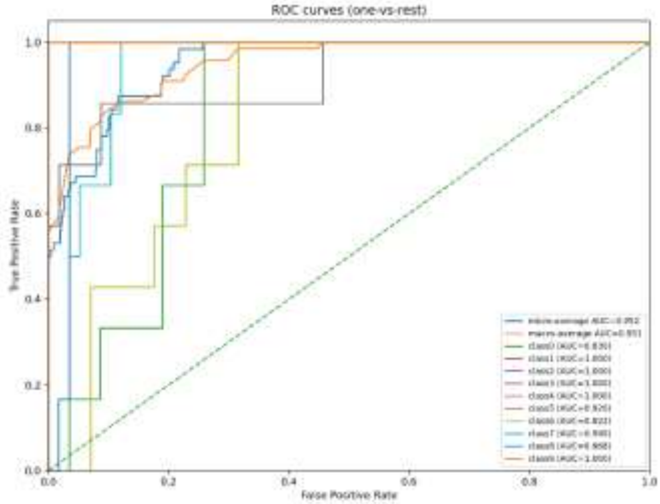


Fig-5: ROC curves- AlexNet(base)model

AlexNet5 (base): Micro/macro AUC ≈ 0.95 (good overall ranking), but class AUCs are poor separability for NSR (≈0.83) and especially RVH (≈0.82); LVH ≈0.92; STEMI ≈0.94; others ≈1.00. The inconsistency—high AUC but terrible confusion—means the scores rank positives fairly well, but the argmax/threshold employed won't transform them into accurate labels (i.e., calibration/decision rule issues).

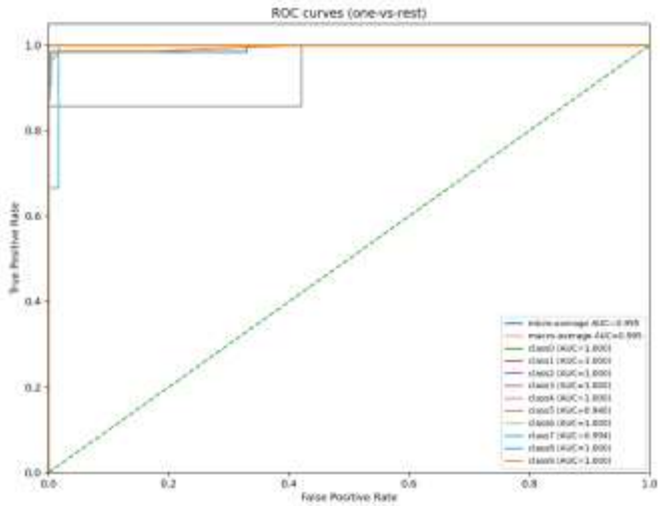
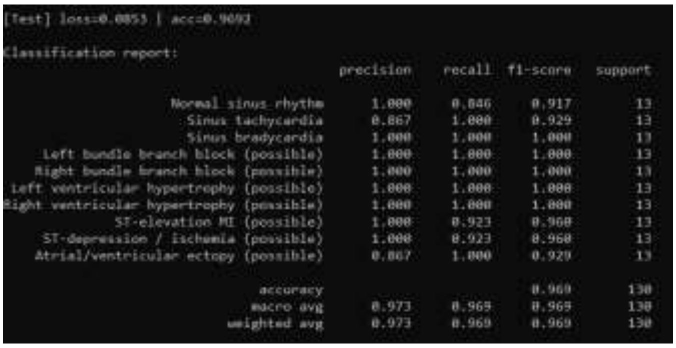


Fig-6: ROC Curves- AlexNet-M-SE model

AlexNet5-M (modified): Micro/macro AUC ≈ 0.995 with per-class AUCs ~1.00 (LVH ≈0.94; STEMI ≈0.994). This shows excellent separability and is as expected with the clean diagonal of the confusion matrix. Bottom line: The BN + Dropout + SE variant of AlexNet substantially enhances class detection for LVH/RVH/ST change and has almost perfect ROC performance. Squeeze the last drop if you must: calibrate probabilities (temperature scaling), re-weight decision thresholds by class (in the context of asymmetric costs), and add targeted data/augmentation for the remaining confusions (brady↔LBBB, LVH↔RVH).

Dataset=440; test_set= 130:

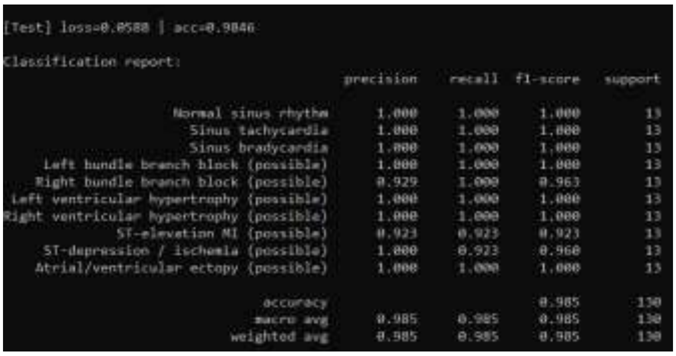
AlexNet Model:



```
[Test] loss=0.0853 | acc=0.9692

Classification report:
                                      precision    recall  f1-score   support

                 Normal sinus rhythm      1.000     0.846     0.917        13
                   Sinus tachycardia      0.867     1.000     0.929        13
                   Sinus bradycardia      1.000     1.000     1.000        13
     Left bundle branch block (possible)  1.000     1.000     1.000        13
    Right bundle branch block (possible)  1.000     1.000     1.000        13
 Left ventricular hypertrophy (possible)  1.000     1.000     1.000        13
Right ventricular hypertrophy (possible)  1.000     1.000     1.000        13
              ST-elevation MI (possible)  1.000     0.923     0.960        13
      ST-depression / ischemia (possible)  1.000     0.923     0.960        13
     Atrial/ventricular ectopy (possible)  0.867     1.000     0.929        13

                            accuracy                          0.969       130
                           macro avg      0.973     0.969     0.969       130
                        weighted avg      0.973     0.969     0.969       130
```

Fig-7: Classification report- AlexNet(base)model

AlexNet-M model:



```
[Test] loss=0.0588 | acc=0.9846

Classification report:
                                      precision    recall  f1-score   support

                 Normal sinus rhythm      1.000     1.000     1.000        13
                   Sinus tachycardia      1.000     1.000     1.000        13
                   Sinus bradycardia      1.000     1.000     1.000        13
     Left bundle branch block (possible)  1.000     1.000     1.000        13
    Right bundle branch block (possible)  0.929     1.000     0.963        13
 Left ventricular hypertrophy (possible)  1.000     1.000     1.000        13
Right ventricular hypertrophy (possible)  1.000     1.000     1.000        13
              ST-elevation MI (possible)  0.923     0.923     0.923        13
      ST-depression / ischemia (possible)  1.000     0.923     0.960        13
     Atrial/ventricular ectopy (possible)  1.000     1.000     1.000        13

                            accuracy                          0.985       130
                           macro avg      0.985     0.985     0.985       130
                        weighted avg      0.985     0.985     0.985       130
```

Fig-8: Classification report- AlexNet-M-SE model

The above implementation shows that, the dataset hs been increased, so the test_set. The modified AlexNet-M showed an accuracy of 98.5% while the base model showed only 96.9% accuracy. These results are far better compared to previous dataset results. The base model and modified models work better with increasing the samples in datasets.

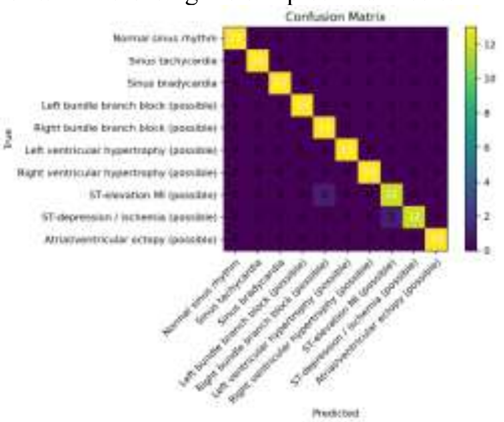


Fig-9: ConfusionMatrix- AlexNet(base)model

Confusion matrix: total errors = 4 → 126/130 = 96.9% accurate. Error modes: Normal sinus rhythm → Sinus tachycardia (2/13) and the same ST-elevation MI ↔ ST-depression swap (1 each); the rest of the classes are perfect. This indicates that the base model overcalls rate-related changes occasionally and swaps the two ST syndromes minimally. 13 test samples per class; 2 total errors → 128/130 = 98.5% accuracy. The errors are a small, clinically reasonable spill: ST-elevation MI ↔ RVH (1 case) and ST-elevation MI ↔ ST-depression (1 case). All other classes are on the diagonal precisely. Practical implication: residual ambiguity is limited to the ST/RVH neighborhood; per-class recall/precision for the affected classes ≈ 12/13 ≈ 92%, others ≈ 100%.

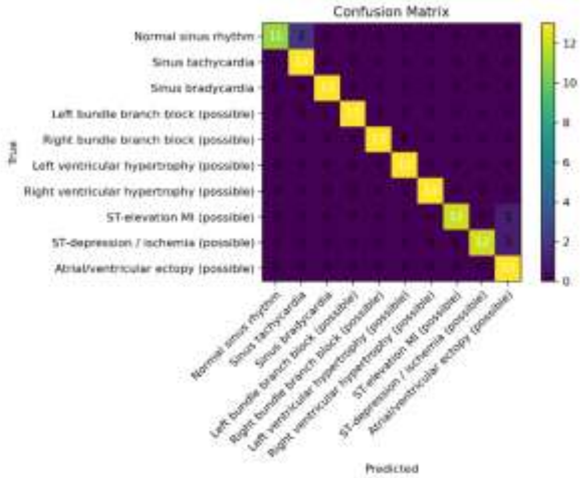


Fig-10: ConfusionMatrix- AlexNet-M-SE model

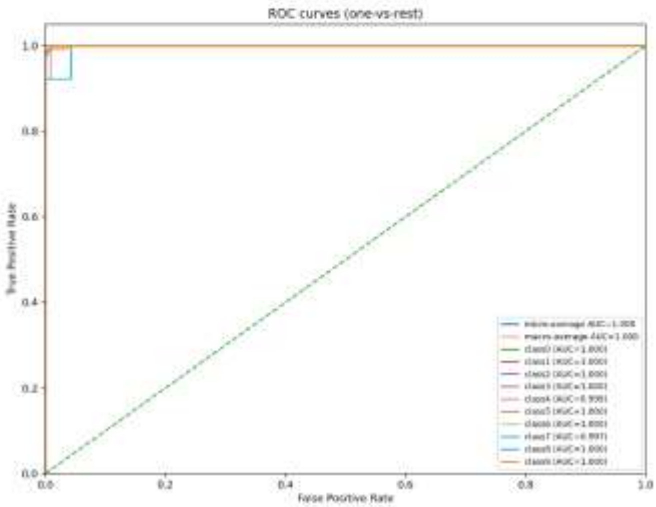


Fig-11: ROC curves- AlexNet(base)model

Micro/macro AUC ≈ 1.000 too; per-class AUCs are ~1.00 with small drops (STEMI ~0.999, ST-depression ~0.998). Ranking is great, but decision boundaries cause some mislabels.

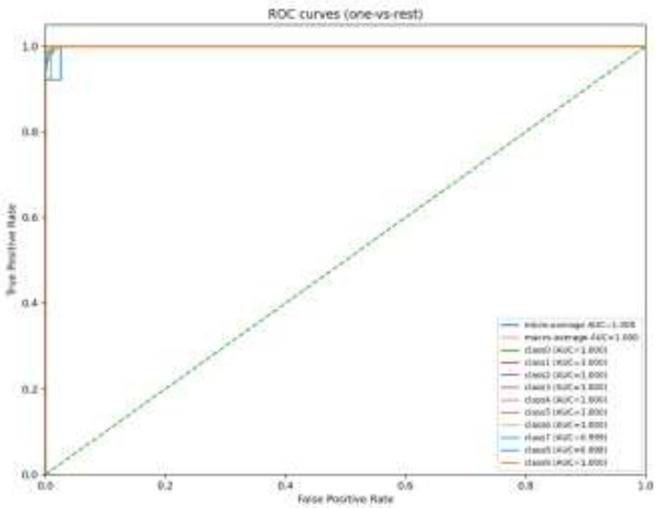


Fig-12: ROC Curves- AlexNet-M-SE model

Micro/macro AUC = 1.000; individual AUCs are ~1.00 with the worst being ~0.997 (STEMI) and ~0.999 (RBBB). So separability is basically ideal; the label mistakes are threshold/argmax effects and not bad ranking.

Bottom line: Both models distinguish classes very well, but the modified i.e., AlexNet-M-SE is demonstrably better—higher accuracy and fewer, more localized mistakes. If you need to squeeze out the final mistakes, attempt per-class threshold tuning or probability calibration, and include targeted augmentation/examples for STEMI versus ST-depression (and ST-like changes adjacent RVH).

4. CONCLUSIONS

In brief, both models classify ECG exceptionally well on this data set, but the modified AlexNet-M-SE is clearly superior. AlexNet5-M-SE (BatchNorm + Dropout + SE) achieves 128/130 correct (≈98.5% accurate) with only two clinically plausible slips—one ST-elevation MI misclassified as RVH and one as ST-depression—while all other classes are classified correctly. By comparison, the vanilla AlexNet5 achieves 126/130 correct (≈96.9% accurate) and has two additional errors: Normal sinus rhythm → Sinus tachycardia and the same small STEMI ↔ ST-depression exchange. ROC curves corroborate the image: both models exhibit near-perfect separability with micro/macro AUC ≈ 1.000, and classwise AUCs ≈0.998–1.000, i.e., score ranking is extremely resilient; the handful of errors stem from decision boundaries, not weak feature acquisition. Clinically, the adapted model's tighter confusion matrix reflects superior generalization to rate-related and ST-segment pathologies, while the vanilla model overcalls sinus tachycardia and slightly blurs the two ischemic patterns. External validation on larger, heterogeneous cohorts is recommended before clinical deployment.

ACKNOWLEDGEMENT

This is to acknowledge my mentor Dr. Manjula Devarakonda Venkata for making this work success. My sincere thanks to management of Pragati Engineering College(A) for giving me constant support.

REFERENCES

[1] Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” Proceedings of the 37th International Conference on Machine Learning (ICML 2020), PMLR, vol. 119, 2020, pp. 1597–1607.
[2] Deng, Jia, et al. “ImageNet: A Large-Scale Hierarchical Image Database.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.
[3] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[4] Howard, Andrew G., et al. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.” arXiv preprint arXiv:1704.04861, 2017.
[5] Ioffe, Sergey, and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), 2015, pp. 448–456.
[6] Kingma, Diederik P., and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” International Conference on Learning Representations (ICLR 2015), 2015.
[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.
[8] Nair, Vinod, and Geoffrey E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines.” Proceedings of the 27th International Conference on Machine Learning (ICML 2010), 2010, pp. 807–814.
[9] Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv preprint arXiv:1409.1556, 2014.
[10] Goldman, M. J. Principles of Clinical Electrocardiography. 1970
[11] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, “Deep learning applications in ECG analysis and disease detection: An investigation study of recent advances,” IEEE Access, vol. 12, pp. 126258–126284, Jan. 2024, doi: 10.1109/access.2024.3447096.
[12] Zheng, Jianwei, et al. "A large scale 12-lead electrocardiogram database for arrhythmia study" (version 1.0.0). *PhysioNet* (2022). RRID:SCR_007345. https://doi.org/10.13026/wgex-er52
[13] Zheng, J., Chu, H., Struppa, D. *et al.* Optimal Multi-Stage Arrhythmia Classification Approach. *Sci Rep* **10**, 2898 (2020). https://doi.org/10.1038/s41598-020-59821-7
[14] Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220." (2000). RRID:SCR_007345.
[15] R. Balestriero and R. Baraniuk, “Batch Normalization Explained,” in *Proc. International Conference on Learning Representations (ICLR)*, 2023.
[16] I. Salehin and D.-K. Kang, “A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain,” *Electronics*, vol. 12, no. 14, art. 3106, Jul. 2023, doi: 10.3390/electronics12143106.
[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.