International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 11 [November 2026] PP: 221-230

Explainable Alzheimer's Disease Classification from Brain MRI Using EfficientNetB0, Grad-CAM and LIME: A Software Prototype Study

Syed Khubayb Ur Rahman, Shankar, Nagaraj, ShohanPal
Department of CSE, APSCE
Corresponding author: *Syed Khubayb Ur Rahman*
Academic guide: *Prof. Parashivamurthy C*

ABSTRACT: *ThispaperreportsasoftwareprototypestudyofanEfficientNetB0-basedbrainMRIclassification application integrating Grad-CAM, LIME and persistent PDF reporting. The objective is to verify the implemented workflow and correspondence between an uploaded image, its model outputs and its exported record. The application prepares an image, produces scores for four dataset categories and stores the input, explanations and metadata under a unique identifier. The recorded functional evaluation covered 13 saved prediction records and four upload-to-export trials. All tested records passed the specified score-consistency, asset-presenceand PDF-exportchecks;each trialproduced six imageassetsand atwo-pagereport.Asupporting inventory of 44,000 image files identified 43,538 distinct SHA-256 hashes and 462 redundant files. Evaluation unitswere softwarerecordsand requests,notindependentlyrecruitedpatients.Historical notebook classification figures are discussed only as unverified development artifacts. The findings support operation of the tested prototype and preservation of prediction evidence. This study does not establish diagnostic accuracy, clinical validity, patient-level generalization or explanation fidelity.*

KEYWORDS: *Software prototype; functional evaluation; brain MRI; EfficientNetB0; Grad-CAM; LIME; persistent reporting.*

Date of Submission: 09-11-2026 Date of acceptance: 22-11-2026

I. INTRODUCTION

Brain MRI classification projects often combine model development with a web interface intended to demonstratepredictionsonindividualimages.Forsuchaprototype,ausabledemonstrationrequiresmorethana predicted class. The displayed input, class scores, explanation images and exported document must all refer to the same inference event. If these outputs are overwritten by later uploads, the user cannot reliably recover the evidence associated with an earlier result. Preserving this association is therefore an engineering requirement that can be evaluated independently of clinical performance.

MRI classification for Alzheimer's disease also requires careful interpretation of the underlying evidence. Wen et al. [1] showed how differences in participant selection and evaluation procedures complicate comparisons between published classifiers. Their reproducible evaluation highlights the importance of avoiding data leakage. These concerns are particularly relevant when a collection contains two-dimensional images withoutan accompanyingpatientregistry:acountoffilesdoesnotestablish acountofindependentparticipants.

This is a software prototype development and functional evaluation study. Its research question is whether a local application consistently associates one uploaded image with its scores, Grad-CAM and LIME artifacts, saved metadata and exported PDF. The contribution is the documented integration and verification of these components. Outcomes concern record consistency, artifact availability, report retrieval and bounded workflow observations. The design does not estimate sensitivity, specificity or diagnostic accuracy in a clinical population.

Partsof the application codewere adaptedfrom thepublicAlzheimer'sdisease predictionrepositoryby Likith Pentakota and Learn Alloid, which already describes MRI classification, Grad-CAM and a Flask application. The reused code and existing classification, Grad-CAM and Flask workflow are credited to that project and are not presented as original inventions of this study. This paper examines the current local implementation, including its persistent report organization, foreground-restricted explanation displays and reproducibility audit. The available saved artifact, notebook and prediction records define the evidenceexamined here.

The study treats the application as a software artifact whose observable behavior can be checkedagainst stated requirements. This framing draws on design-science work by Hevner et al. [2] and the separation of development, demonstration and evaluation described by Peffers et al. [3]. The present evidence concerns a bounded implementation; it does not establish a new learning algorithm or the clinical utility of the artifact.

II. RELATEDWORK

EfficientNet scales convolutional-network depth, width and input resolution jointly [4]. The saved artifact in this prototype contains an EfficientNetB0 feature extractor with a compact classification head. The present work adapts an existing architecture to four image categories and evaluates its integration with explanations and reporting, without proposing a new neural architecture.

Grad-CAM uses class-related gradients at a convolutional layer to construct a coarse spatial attribution map [5]. LIME instead fits an interpretable local approximation using perturbed inputs [6]. The two methods exposedifferentaspectsofaprediction,butagreementbetweentheirvisualizationsisnotproofofapathological finding. Adebayo et al. [7] demonstrate why visual appeal alone is an inadequate way to validate saliency methods.

TheCLAIM2024guidelinepromotestransparentreportingofmedicalimagingAI[8].Wedescribethe evaluation units, model identity and limitations explicitly, while retaining the distinction between software verification and a clinical prediction study.

Ghassemi et al. [9] caution that a plausible explanation display cannot substitute for validation in health care. In this prototype, Grad-CAM and LIME are outputs whose generation and storage can be checked, while their clinical interpretation remains untested. This distinction determines which claims the results can support.

Reproducible machine-learning research requires enough information to connect a reported result to its code, data and experimental conditions [10]. Dataset datasheets [11] and model cards [12] offer complementary ways to record provenance, intended use and evaluation limits. The local audit records the checkpoint identity and software behavior, but the missing patient mapping and test manifest prevent an equivalent reconstructionof diagnostic performance.

III. SYSTEMDESIGNANDDATAAUDIT

ALocalimagecollectionandunitsofanalysis

The local collection contains four directories whose labels are NonDemented, VeryMildDemented, MildDemented and ModerateDemented. Table 1 gives the file inventory. These names are used as dataset classes; their presence does not independently verify disease diagnosis or severity. The inventory counts stored files, including any derived images, and must not be interpreted as a patient cohort. A file was assigned to an exact-duplicate group when its SHA-256 hash matched another file. This audit detects byte-identical copies but does not detect all transformed copies, adjacent slices or multiple scans from one participant.

Table1Localimagefile inventory

Classfolder	Files	Filesnamedaug prefix
MildDemented	10000	144
ModerateDemented	10000	3472
NonDemented	12800	0
VeryMildDemented	11200	0
Total	44000	3616

The project team identified the Kaggle resource under aryansinghal10 [13] as the collection source. Its downloaded version, patient identifiers, acquisition records and source-to-derived-image mapping were unavailable for this evaluation. The aug_ prefix was counted as a filename property; it does not establish a complete augmentation history. The collection is therefore used for a file audit and software demonstrations, without deriving a diagnostic performance estimate.

B Verifiedmodelarchitecture

The saved Keras graph accepts an input tensor of shape 224 by 224 by 3 and contains a frozen EfficientNetB0backbone.Globalaveragepoolingisfollowedbybatchnormalization,dropoutatrate0.4,a256-unit ReLU layer and four softmax outputs. Inspection records 4,383,655 total parameters and 331,524 trainable parameters. The filename models/efficientnetb0_model.keras and the application labels identify this same architecture.

C EfficientNetB0trainingconfiguration

The EfficientNetB0 notebook defines a frozen backbone and trains the classification head using Adam with learning rate 0.001 and sparse categorical cross-entropy. It specifies ImageNet initialization, batch size 32 andamaximumof40epochs,withearlystopping,learning-ratereductionandcheckpointsaving.Thesesettings describe the retained training code, not a verified completed training history for the deployed weights. The present application study uses the existing checkpoint without retraining.

D Imagepreparationandinference

An uploaded image is decoded as RGB. A grayscale threshold of 10 is used to find external contours, and the bounding rectangle of the largest contour defines the crop. The crop is resized to 224 by 224 pixels. Ifno contour is found, the original image is retained before resizing. Keras documents embedded preprocessingfor EfficientNet applications [14]. Output indices 0, 1, 2 and 3 correspond respectively to NonDemented, VeryMildDemented, MildDemented and ModerateDemented. The class with the largest output score suppliesthe displayed label. RGB arrays are passed as float32 values in the 0-255 range because rescaling is embeddedin the EfficientNetB0 graph. Direct inference and LIME perturbations use the same preprocessing helper.

The displayed confidence is the largest softmax score multiplied by 100 and rounded to two decimal places. It is an uncalibrated model score. Guo et al. [15] distinguish confidence calibration from classification performance; no calibration experiment was conducted here. The application does not estimate predictive uncertainty or provide an abstention rule for unfamiliar images. In particular, a high softmax value is not reported here as the probability that a patient has Alzheimer's disease.

E Visualexplanationsandforegroundtreatment

For Grad-CAM, the application locates the last convolutional layer, including layers inside a nested backbone. It differentiates the selected softmax output with respect to that layer's activations, spatially averages the gradients, combines the feature channels with those weights and retains positive contributions. The resulting map is normalized and resized for display. For the colored overlay, values within the foreground are further scaled using the fifth and ninety-ninth percentiles; therefore, overlay intensity is a relative display quantity and cannot be compared directly across images. A separate foreground mask restricts the colored overlay: grayscale values above 18 are morphologically opened, the largest connected component is retained, and one erosion step removes part of the boundary. This is an intensity-based display mask and is not a validated anatomical segmentation.

For LIME, the implementation uses a fixed random seed of 42, 100 perturbation samples and masked SLIC segmentation [16] with a target of 65 segments, compactness 12 and zero smoothing. The prediction wrapper restores all background pixels before inference on each perturbation. The display highlights up to five tissue-containing segments ordered by the absolute magnitude of their local coefficients, with supporting and opposing coefficients shown in different colors. These segments are image regions rather than anatomically registered brain structures. The foreground restriction constrains the visualization and perturbation procedure; it does not demonstrate that the original classifier ignores background information.

F PersistentrecordsandPDFgeneration

Eachsubmissionreceivesauniqueidentifierandaseparatestoragedirectory.Theapplicationsavesthe original image, class-score chart, raw Grad-CAM heatmap, Grad-CAM overlay, LIME overlay and LIME coefficient chart. A JSON manifest records the identifier, timestamp, input filename, predicted class, all four scores and flags indicating available assets. The manifest is written to a temporary file and then replaced atomically. Artifact production is serialized within the process because plotting uses shared state.

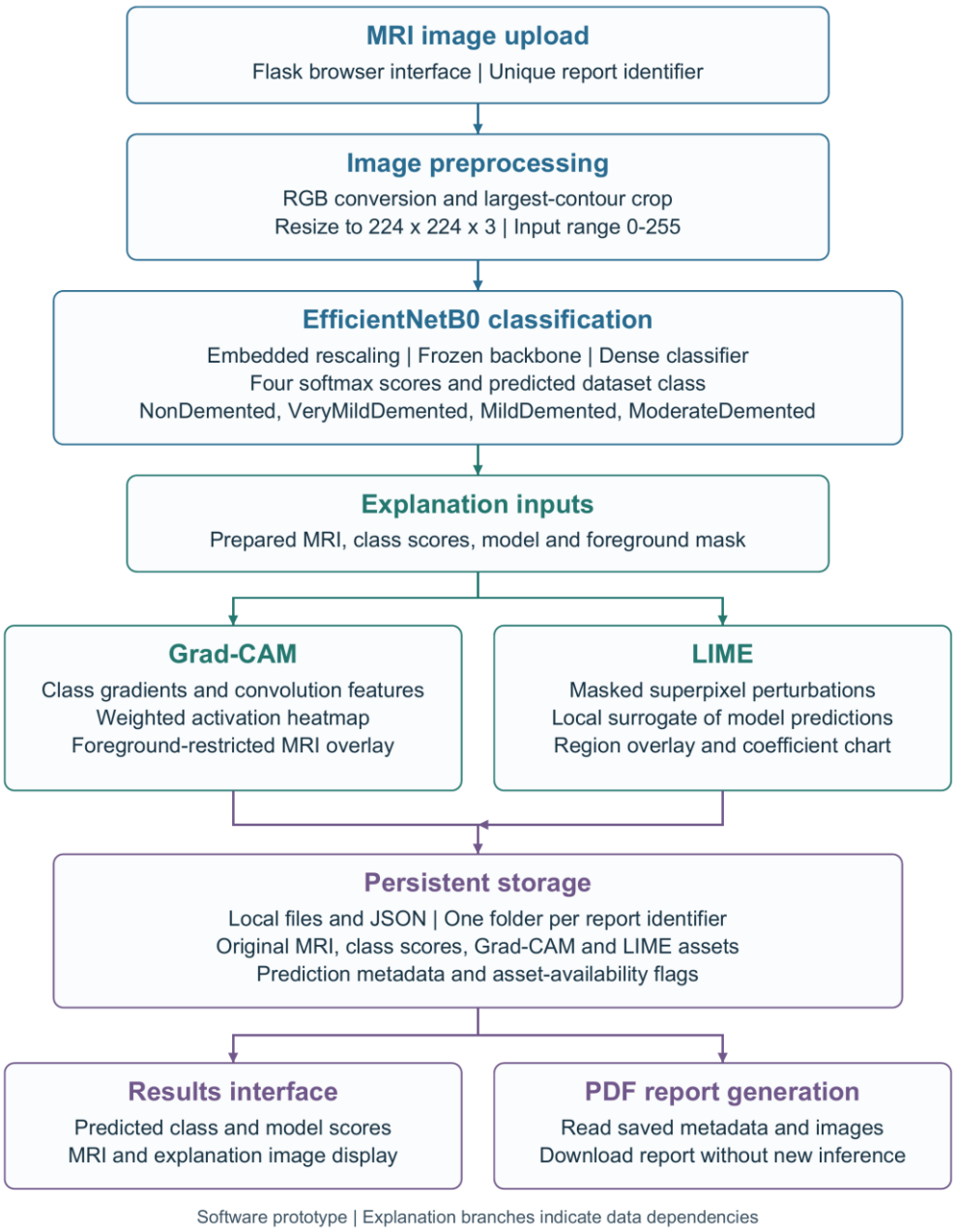

Figure 1 System architecture of the EfficientNetB0 prototype

The download route constructs the PDF from this saved record and its associated images. It does not rerun the neural classifier or regenerate the explanations. Report identifiers are restricted to 32 lowercase hexadecimal characters, and image requests are limited to an asset allowlist. Missing assets produce explicit fallback text in the exported document. These mechanisms support record consistency; they do not constitute a complete authentication, privacy or deployment-security design.

IV. SOFTWAREVERIFICATIONMETHODS

The design is a single-application software prototype study using an existing checkpoint. The case boundary, evaluation units and limits on generalization are reported explicitly, drawing on software-engineering case-study guidance [17]. The study containsno recruited patient cohort, controlled clinical comparison or user-acceptance experiment. Its evidence comes from file inspection, saved-record checks and four application requests.
Thecheckedrequirementswerethatarecordretainitsownidentifier,thatitsdisplayedlabelagreewith its stored scores, that its expected images remain retrievable, and that PDF export use the saved record. Each check produced a pass or failure for the relevant record or request. The selection rule and denominators below make these observations inspectable, but no statistical inference is made from the small request sample.

The evaluation was performed in a single local Windows CPU session. The application environment used Python 3.13.12, TensorFlow 2.21.0, Keras 3.15.1, Flask 3.1.3, LIME 0.2.0.1 and scikit-image 0.26.0. The saved model was fingerprinted with SHA-256 to link the tests to a particular weight artifact. Test inputs, generated manifests and machine-readable outcomes were retained separately from the original prediction records.

The first evaluation unit was the complete set of 13 saved prediction manifests present when the audit ran. For each record, the checks compared the stored class with the largest class score, compared the displayed percentage with the rounded maximum score, checked all six asset files, and generated a PDF. Export checks verified a two-page document containing the report identifier and predicted label. Five malformed identifiersand two disallowed asset names were tested separately. The missing-image condition was exercised byexporting a record from a directory without its image assets.

The second unit was a fresh application request. One image was selected from each class folder by takingthefirstsupportedimagefilenameinsortedorder.EachimagewassubmittedthroughtheFlasktestclient to the prediction route, followed by a request to the report download route. The checks covered the response status, manifest creation, six available assets, report length and identifier preservation. The Grad-CAM overlay was also compared with its input outside the computed mask to verify exact preservationof those pixels. This is a software property of the display operation, not a test of explanation fidelity.

Elapsed time was measured around each prediction-route request using a monotonic timer. It includes inference, explanations, plotting, disk writes and HTML response construction. It excludes model loading, the later PDF download and the additional overlay check. Each image was submitted once, with no concurrent requests. The four observations are descriptive measurements of this session; they are not a throughput benchmark or a latency distribution for deployed users.Training overlap and patientindependence are unknown for these images, so no accuracy statistic is derived from the four trials.

V. PROTOTYPEVERIFICATIONRESULTSANDDISCUSSION

ACorpusandmodelaudit

Hashing the 44,000 files yielded 43,538 distinct hashes. There were 195 groups of byte-identical files, accounting for 462 files beyond one representative per group, or 1.05 percent of the collection. None of these exact-duplicate groups crossed the four folder labels. Across the same collection, 3,616 filenames used the aug_ prefix; these countsmay overlap with the exact-duplicate groups. The absence of cross-label duplicates does not establishindependencebetweentrainingandtestsets.Theauditdoesnotrecoverpatientidentities,source-image relationships or the partition used to train the saved artifact.

The model configuration identifies EfficientNetB0 as the deployed architecture. All four output categories, the RGB input contract and the checkpoint fingerprint are recorded independently of its displaylabel. This links the functional evaluation to a specific saved graph rather than a model name alone.

B Storedrecordsandfaulthandling

Table2Observedfunctionaloutcomes

Check	Passedorobserved	Denominator
Storedlabeland roundedscore agree	13	13records
Allsixstored imagefilespresent	13	13records
Twopageexportwithidentifierand label	13	13exports
Malformedidentifiersrejected	5	5identifiers
Disallowedassetnamesrejected	2	2names
Missingimagefallback textpresent	1	1export
Freshpredictionanddownload succeed	4	4requestseach
Freshsubmissionsproducesixassets	4	4submissions
Overlaybackgroundpixelsunchanged	4	4images

All 13 stored records passed the specified consistency and export checks. The malformed identifiers were rejected by the directory validator, and both disallowed names were rejected by the asset helper. The missing-image export contained the expected fallback text. These bounded cases establish behavior for the inputs tested. They do not demonstrate resistance to all malformed requests, simultaneous users or adversarial uploads.

C Recordedworkflowtrials

Table3Foursinglerequestobservations

Inputfolder	Displayedclass	Maximumscore percent	Requesttimeseconds
MildDemented	MildDemented	95.24	5.57
ModerateDemented	ModerateDemented	100.00	2.88
NonDemented	NonDemented	98.35	3.29
VeryMildDemented	VeryMildDemented	76.23	3.48

Each fresh submission produced the six expected images and a two-page downloadable report containingitsownidentifier.Themeasuredprediction-routetimesrangedfrom2.88to5.57seconds;themedian of the four observations was 3.38 seconds. The first request was slower than the later requests. Initialization or tracing within the first inference may contribute to this difference, but the experiment did not isolate that cause. Each displayed class matched its source folder in these four demonstrations. Because the inputs are neither an independent cohort nor an established held-out test set, this agreement must not be interpreted as 100 percent diagnostic accuracy.

For all four images, every pixel outside the computed foreground mask remained unchanged in the directly checked Grad-CAM overlay. This verifies that color blending respected the mask in the tested display operation. It does not validate the mask as brain tissue, confirm that a highlighted region contains disease, or show that the explanation is faithful to clinically meaningful features. No radiologist assessment, model-randomization test, deletion or insertion experiment, or multi-seed LIME stability analysis was performed.

D Implicationsforaneducationalprototype

Persisting metadata and images together links each exported report to one prediction event. Exporting saved outputs avoids replacing an earlier result with a later inference. The tests establish this behavior for the records examined; they do not quantify improvement over an alternative system.

Visual outputs from the recorded workflow trials

Figure 2 presents the saved input and explanation images for the four requests in Table 3. Each row uses files from the same prediction record. The input panel is the uploaded image; the explanation panels reflect the application crop and resize. These examples illustrate the generated outputs and their association with a prediction record.

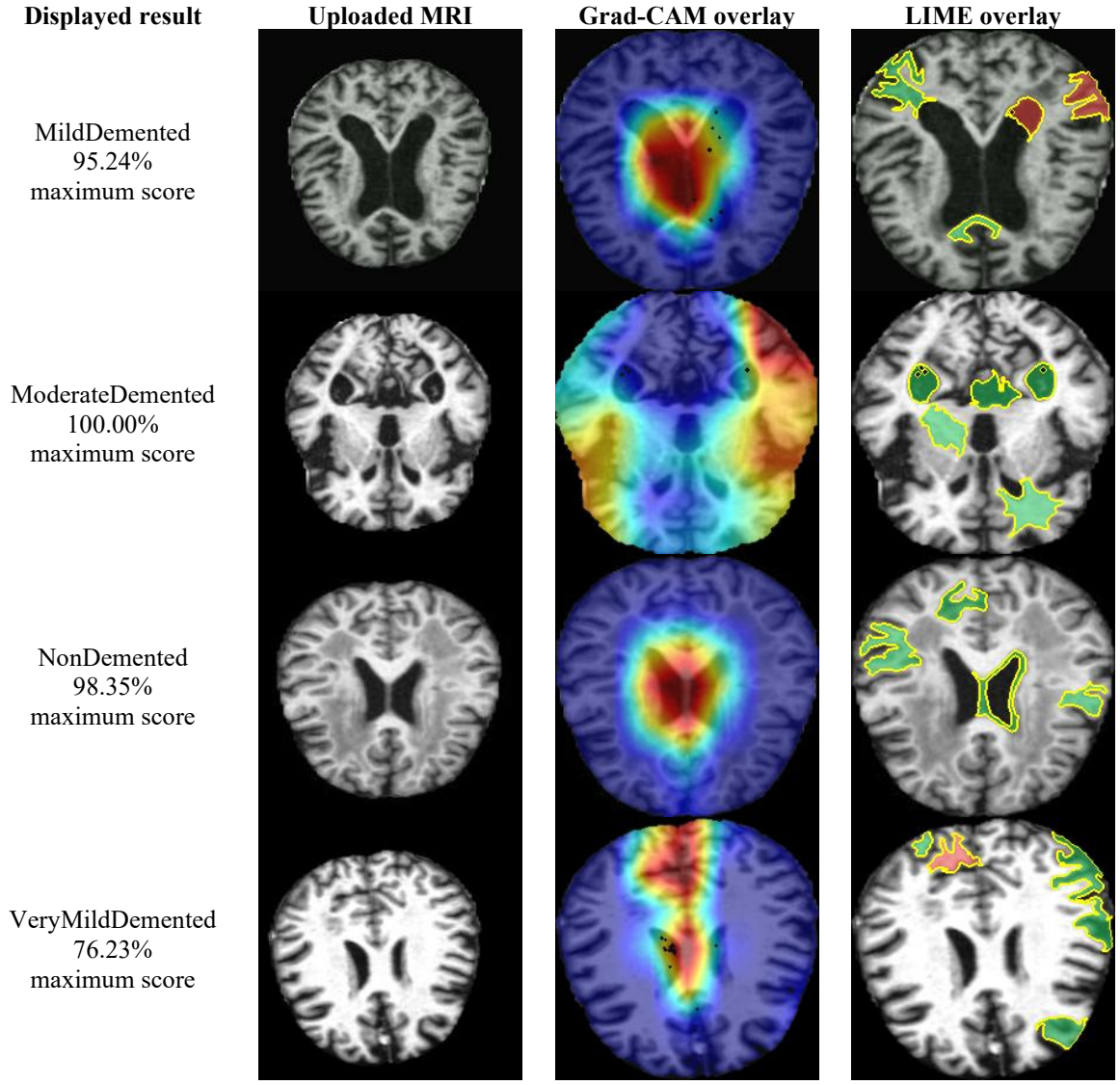

Figure 2 MRI inputs with saved Grad-CAM and LIME outputs

Rows follow Table 3. Scores are uncalibrated model outputs. Highlighted regions are software explanations and have not been validated as disease locations or anatomical segmentations. The four examples do not provide an independent diagnostic-accuracy estimate.

Class score results for the same four requests

Figure 3 plots all four stored class scores for each illustrated request. The selected class is the largest score. The VeryMildDemented example also assigns 23.77% to NonDemented, showing information that would be hidden by reporting only the winning label.

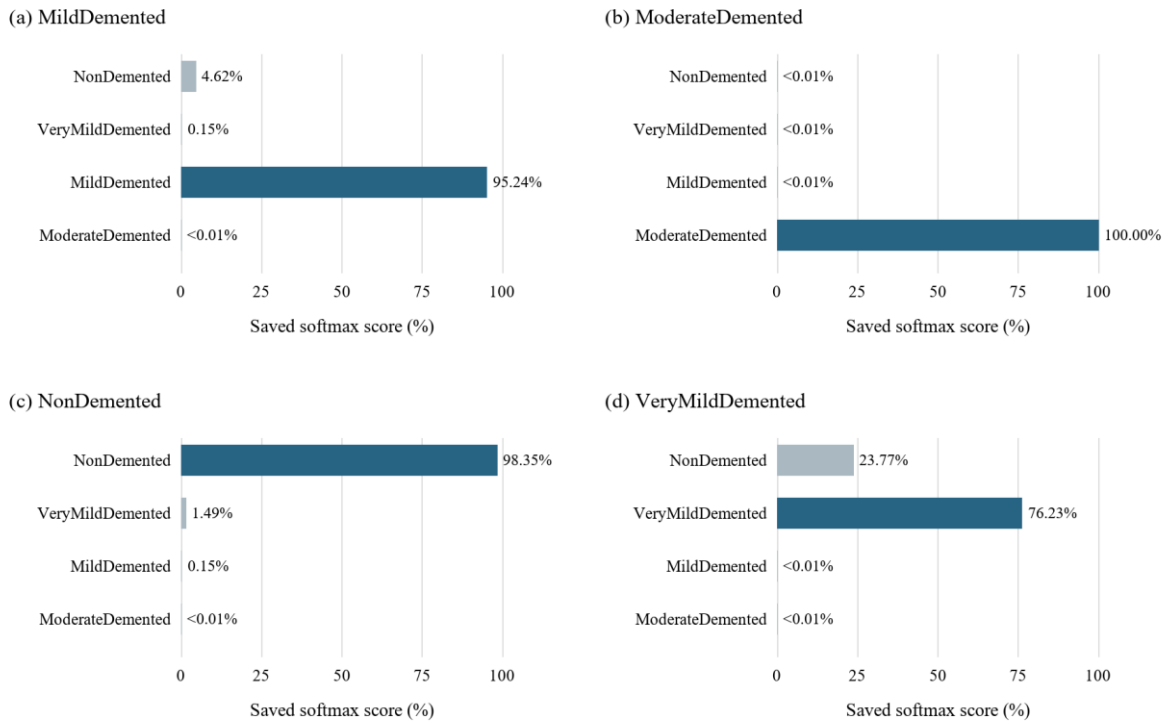

Figure 3 Saved class-score distributions from the four workflow trials

Charts are plotted from the original prediction.json records used for Figure 2 and Table 3. Values are displayed to two decimal places; values below 0.01% are marked accordingly. A displayed value of 100.00% is rounded and does not imply certainty. These plots describe individual software outputs, rather than cohort-level performance.

VI. LIMITATIONSANDFURTHEREVALUATION

Dataset provenance is the principal constraint on interpreting the classifier. The collection lacks a verified patient registry, acquisition protocol and source-to-derived-image mapping. Exact-file hashes identify only one category of duplication. Patient-level splitting, grouping of all slices and augmentations from the same source, and anindependentlysourced evaluation cohort arerequiredto investigate generalization. Thefour class names should remain dataset labels until their clinical derivation and ground truth are documented.

The model-preparation path also differs between notebook and application: the notebook uses resizing with padding, while the application crops and then resizes. The effect of this difference on predictions was not measured. Explanation quality remains unvalidated: the display mask is heuristic, Grad-CAM is coarse, and LIME uses a small perturbation budget without a reported surrogate-fidelity score. Future evaluation should examine preprocessing consistency, uncertainty calibration, explanation sensitivity and expert interpretation before clinical usefulness is assessed.

The software evaluation covers one process on one local computer, 13 stored records and four upload-to-export demonstrations. It does not measure sustained load, concurrent users, calibrated uncertainty or production access controls. The observed request times describe the recorded session and should not be generalized to other hardware or workloads.

Historicaltrainingartifactandprovenancelimits

The notebook retains a four-class EfficientNetB0 confusion matrix with 400 evaluated images, including 100 examples per reference class. Its diagonal contains 84, 80, 87 and 100 correct predictions in the fixed application class order, giving 351 correct classifications and 87.75 percent accuracy. Table 4 reproduces the counts. These counts are retained development artifacts, not outcomes of the prototype verification study ora validated estimate of deployed-model performance.

Table4Historicalnotebookartifactoutsideprototypevalidation

Reference/ predicted	NonDemented	VeryMild	Mild	Moderate
NonDemented	84	12	4	0
VeryMildDemented	13	80	7	0
MildDemented	5	2	87	6
ModerateDemented	0	0	0	100

Rows are reference categories and columns are predicted categories. VeryMild, Mild and Moderate abbreviate VeryMildDemented, MildDemented and ModerateDemented. The exact class recalls calculated from the retained countsare84,80,87 and 100 percent. The macro-average recall is87.75percent because each class has equal support. These figures describe this recorded subset and do not establish patient-independent diagnostic performance.

The splitting code declares 4,400 test images, which differs from the 400-image evaluation support. A frozen test manifest and per-image predictions are unavailable to reconcile these counts or link the result conclusively to the deployed weights. The historical accuracy therefore cannot be treated as an independently reproduced performance estimate for the application or used to claim superiority over other systems.

VII. CONCLUSION

This software prototype study documents the integration of EfficientNetB0, Grad-CAM, LIME and persistent PDF reporting in a local Flask application. The recorded verification covered 13 saved records and four upload-to-export trials. Within those bounded cases, the application preserved record identifiers, score consistency, image assets and report retrieval. These are software functionality findings. They do not constitute evidence of diagnostic accuracy, clinical effectiveness or validated explanations. Historical notebook figures remain development context with unresolved checkpoint and split provenance. A separate patient-independent evaluation, with a frozen test manifest and verified reference labels, would be required to investigate diagnostic performance.

ACKNOWLEDGEMENT

The authors acknowledge their academic guide, Prof. Parashivamurthy C (psmurthycs@gmail.com),for guidance on the project. They also acknowledge Likith Pentakota and Learn Alloid for the public software repository from which parts of the application code were adapted.

DECLARATIONS

Study design. Softwareprototypedevelopment andfunctional evaluation;nodiagnostic accuracystudy was conducted.

Ethics and consent. This study examines an existing MRI image collection and saved software records. Itinvolvednonewparticipantrecruitment,MRIacquisitionorclinicalintervention.Noidentifiableparticipant details are reported. An institutional ethics approval or exemption reference has not been documented for thisstudy.

DATAANDSOFTWAREAVAILABILITY

The dataset resource is cited in [13]. The evaluated modifications, checkpoint fingerprint, verification script and machine-readable outcomes remain in the local project archive. The evaluated checkpoint and audit package have not been publicly released. Dataset version and reuse conditions require confirmation before redistribution.

AIASSISTANCE

AI tools assisted drafting, manuscript organization, reference checking, language editing and formatting. The authors are responsible for the final manuscript and the accuracy of its claims and citations.

REFERENCES

[1] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, et al., "Convolutional neural networks for classification of Alzheimer's disease: Overview andreproducible evaluation," Medical Image Analysis, vol. 63, article 101694, 2020. https://doi.org/10.1016/j.media.2020.101694

[2] A. R. Hevner, S. T. March, J. Park and S. Ram, "Design science in information systems research,"MIS Quarterly, vol. 28, no. 1, pp. 75-105, 2004. https://doi.org/10.2307/25148625

[3] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee, "A design science research methodology for information systemsresearch," Journal ofManagement Information Systems,vol.24,no.3,pp.45-77,2007.https://doi.org/10.2753/MIS0742-1222240302

[4] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," Proceedings of the 36th InternationalConference on Machine Learning, PMLR, vol. 97, pp. 6105-6114, 2019. https://proceedings.mlr.press/v97/tan19a.html

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual explanations from deep networks

viagradient-based localization," Proceedings of the IEEE International Conference on Computer Vision, pp. 618-626, 2017.https://arxiv.org/abs/1610.02391

[6] M. T. Ribeiro, S. Singh and C. Guestrin, "Why Should I Trust You?: Explaining the predictions of any classifier," Proceedings of the22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, 2016.https://arxiv.org/abs/1602.04938

[7] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt and B. Kim, "Sanity checks for saliency maps," Advances in NeuralInformation Processing Systems, vol. 31, 2018. https://arxiv.org/abs/1810.03292

[8] A.S.Tejani,M.E.Klontzas,A.A.Gatti,etal.,"ChecklistforArtificialIntelligenceinMedicalImaging(CLAIM):2024Update," Radiology:ArtificialIntelligence,vol.6,no.4,articlee240300,2024.https://doi.org/10.1148/ryai.240300

[9] M. Ghassemi, L. Oakden-Rayner and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in healthcare," The Lancet Digital Health, vol. 3, no. 11, pp. e745-e750, 2021. https://doi.org/10.1016/S2589-7500(21)00208-9

[10] J. Pineau, P. Vincent-Lamarre, K. Sinha, et al., "Improving reproducibility in machine learning research (a report from the NeurIPS2019 reproducibility program)," Journal of Machine Learning Research, vol. 22, no. 164, pp. 1-20, 2021.https://jmlr.org/papers/v22/20-303.html

[11] T.Gebru,J.Morgenstern,B.Vecchione,J.W.Vaughan,H.Wallach,H.DaumeIIIandK.Crawford,"Datasheetsfordatasets," CommunicationsoftheACM,vol.64,no.12,pp.86-92,2021.https://doi.org/10.1145/3458723

[12] M. Mitchell, S. Wu, A. Zaldivar, et al., "Model cards for model reporting," Proceedingsof the Conference on Fairness, Accountability,and Transparency, pp. 220-229, 2019. https://doi.org/10.1145/3287560.3287596

[13] aryansinghal10, "Alzheimer's Disease Multiclass Images Dataset," Kaggle dataset, accessed November 1, 2026.https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented

[14] Keras, "EfficientNet models," Keras Applications documentation, accessed November 1, 2026.https://keras.io/api/applications/efficientnet/efficientnet_models/

[15] C. Guo, G. Pleiss, Y. Sun and K. Q. Weinberger, "On calibration of modern neural networks," Proceedings of the 34th InternationalConference on Machine Learning, PMLR, vol. 70, pp. 1321-1330, 2017. https://proceedings.mlr.press/v70/guo17a.html

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixelmethods," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274-2282, 2012.https://doi.org/10.1109/TPAMI.2012.120

[17] P. Runeson and M. Host, "Guidelines for conducting and reporting case study research in software engineering," Empirical SoftwareEngineering, vol. 14, pp. 131-164, 2009. https://doi.org/10.1007/s10664-008-9102-8