The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

Interpretable Brain Tumor Classification using Vision Transformers with Attention Maps

Dr Humera Shaziya
Email-humera.shaziya@gmail.com
Department of Informatics, Nizam College, Osmania University

అటెన్షన్ మ్యాప్‌లతో విజన్ ట్రాన్స్‌ఫార్మర్‌లను ఉపయోగించి అర్థమయ్యే మెదడు కణితి వర్గీకరణ

డాక్టర్ హుమేరా షాజియా
ఈమెయిల్ - humera.shaziya@gmail.com
ఇన్ఫర్మేటిక్స్ విభాగం, నిజాం కళాశాల, ఉస్మానియా విశ్వవిద్యాలయం

వియుక్త

ViT) ఉపయోగించి మెదడు కణితి వర్గీకరణ కోసం అర్థమయ్యే లోతైన అభ్యాస విధానాన్ని అందిస్తుంది . ఈ అధ్యయనం వర్గీకరణ పనితీరును మాత్రమే కాకుండా, శ్రద్ధ పటాలను దృశ్యమానం చేయడం ద్వారా వివరణాత్మకతను కూడా నొక్కి చెబుతుంది. అసమతుల్య తరగతులతో కూడిన మెదడు MRI డేటాసెట్ ఉపయోగించబడింది మరియు సాధారణీకరణను మెరుగుపరచడానికి డేటా పెంపుదల వర్తించబడింది. ప్రతిపాదిత ViT- ఆధారిత నమూనాకు ఖచ్చితత్వం, ఖచ్చితత్వం, రీకాల్, F1-స్కోర్ మరియు AUC/ROC వంటి ప్రామాణిక కొలమానాలను ఉపయోగించి శిక్షణ ఇవ్వబడింది మరియు మూల్యాంకనం చేయబడింది. ఇంకా, శ్రద్ధ పటాల విజువలైజేషన్ మోడల్ కణితి ప్రాంతాలపై ఎలా దృష్టి పెడుతుందో, విశ్వసనీయత మరియు క్లినికల్ వినియోగాన్ని ఎలా పెంచుతుందో చూపిస్తుంది. ప్రయోగాత్మక ఫలితాలు ViT ఆర్కిటెక్చర్ 94% ఖచ్చితత్వంతో అధిక వర్గీకరణ పనితీరును సాధిస్తుందని, వైద్య ఇమేజింగ్‌లో నిర్ణయం తీసుకోవడానికి మద్దతు ఇచ్చే అర్థమయ్యే అవుట్‌పుట్‌లను అందిస్తుందని సూచిస్తున్నాయి, అయితే CNN బేస్‌లైన్ మోడల్ 80% ఖచ్చితత్వాన్ని సాధిస్తుంది.

కీలకపదాలు

బ్రెయిన్ ట్యూమర్ వర్గీకరణ, విజన్ ట్రాన్స్‌ఫార్మర్, వివరించదగిన AI, అటెన్షన్ మ్యాప్స్, మెడికల్ ఇమేజింగ్, MRI

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

పరిచయం

మెదడు కణితులు ప్రాణాంతక పరిస్థితులు, వీటికి సకాలంలో రోగ నిర్ధారణ మరియు సమర్థవంతమైన చికిత్స అవసరం. మెడికల్ ఇమేజింగ్, ముఖ్యంగా MRI స్కాన్లు, ముందస్తుగా గుర్తించడంలో కీలక పాత్ర పోషిస్తాయి. సాంప్రదాయ యంత్ర అభ్యాస పద్ధతులు చేతితో తయారు చేసిన లక్షణాలపై ఎక్కువగా ఆధారపడతాయి, విభిన్న కణితి నమూనాలను సాధారణీకరించే సామర్థ్యాన్ని పరిమితం చేస్తాయి. లోతైన అభ్యాసంలో ఇటీవలి పురోగతులు, ముఖ్యంగా కన్వల్యూషనల్ న్యూరల్ నెట్‌వర్క్‌లు (CNNలు), వర్గీకరణ ఖచ్చితత్వాన్ని గణనీయంగా మెరుగుపరిచాయి. అయితే, CNNలు తరచుగా వివరణాత్మకతను కలిగి ఉండవు, ఇది క్లినికల్ అప్లికేషన్లలో కీలకమైన అవసరం. విజన్ ట్రాన్స్‌ఫార్మర్లు (ViTలు) CNNలకు శక్తివంతమైన ప్రత్యామ్నాయాలుగా ఉద్భవించాయి, చిత్రాలలో ప్రపంచ ఆధారపడటాన్ని సంగ్రహించడానికి స్వీయ-శ్రద్ధ విధానాలను ఉపయోగిస్తాయి. ముఖ్యంగా, ViT లు ఉత్పత్తి చేసే శ్రద్ధ పటాలు మోడల్ నిర్ణయాన్ని ప్రభావితం చేసిన ఇమేజ్ ప్రాంతాలను చూపించడం ద్వారా వివరించదగిన AI (XAI) కోసం ఒక మార్గాన్ని అందిస్తాయి. ఈ అధ్యయనంలో, మేము ViTలను ఉపయోగించి అర్థం చేసుకోగల మెదడు కణితి వర్గీకరణ నమూనాను ప్రతిపాదిస్తాము మరియు అంచనాల విశ్వసనీయతను అంచనా వేయడానికి దాని శ్రద్ధ పటాలను దృశ్యమానం చేస్తాము.

పరిశోధన లక్ష్యాలు మరియు పద్ధతి

ప్రతిపాదిత పద్ధతి MRI స్కాన్‌లను ఉపయోగించి మెదడు కణితి వర్గీకరణ కోసం ఒక నిర్మాణాత్మక విధానాన్ని అవలంబిస్తుంది. డేటాసెట్ బహుళ కణితి తరగతులను కలిగి ఉన్న మెదడు MRI చిత్రాలను కలిగి ఉంటుంది, తరగతి పంపిణీలో అంతర్లీన అసమతుల్యత ఉంటుంది. దీనిని పరిష్కరించడానికి, డేటా ప్రీప్రాసెసింగ్ నిర్వహించబడింది . ఉపయోగించిన కోర్ మోడల్ ఆర్కిటెక్చర్ ఫైన్-ట్యూన్డ్ విజన్ ట్రాన్స్‌ఫార్మర్ (ViT). ప్రిడిక్టివ్ సామర్థ్యం యొక్క సమగ్ర అవగాహనను నిర్ధారించడానికి మోడల్ పనితీరును మూల్యాంకనం చేశారు. ఇంకా, శిక్షణ పొందిన ViT నుండి అటెన్షన్ మ్యాప్‌లను సంగ్రహించడం ద్వారా ఇంటర్‌ప్రెటబిలిటీని ఫ్రేమ్‌వర్క్‌లో విలీనం చేశారు , తద్వారా మోడల్ యొక్క ఫోకస్ ప్రాంతాల విజువలైజేషన్‌ను అనుమతిస్తుంది మరియు నిర్ణయం తీసుకునే ప్రక్రియలో వివరించదగిన అంతర్దృష్టులను అందిస్తుంది.

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

2. సాహిత్య సర్వే

MRI మెదడు కణితి చిత్రాల బహుళ-తరగతి వర్గీకరణ కోసం హైబ్రిడ్ విజన్ ట్రాన్స్‌ఫార్మర్ (ViT) మరియు డీప్ CNN -ఆధారిత నమూనాలను ప్రతిపాదిస్తుంది. ఇది మోడల్ పారదర్శకత మరియు క్లినికల్ నమ్మకాన్ని పెంచడానికి వివరణాత్మక AI (XAI) పద్ధతులను కూడా అనుసంధానిస్తుంది. మెదడు కణితి గుర్తింపు మరియు వర్గీకరణలో విజన్ ట్రాన్స్‌ఫార్మర్లు (ViT లు) మరియు వివరణాత్మక AI (XAI) వాడకంపై సమగ్ర సర్వేను Paper [2] అందిస్తుంది . ఖచ్చితమైన మెదడు కణితి గుర్తింపు కోసం శ్రద్ధ విధానాలను కలుపుకొని ఆప్టిమైజ్ చేయబడిన హైబ్రిడ్ డీప్ లెర్నింగ్ విధానాన్ని ఈ పత్రం [3] పరిచయం చేస్తుంది . మెదడు కణితి వర్గీకరణ కోసం ద్వి-స్థాయి రూటింగ్ దృష్టిని ఉపయోగించే ద్వి-పూర్వ-ఆధారిత హైబ్రిడ్ మోడల్ బయోట్రాన్స్‌ఎక్స్‌ను పేపర్ [4] ప్రతిపాదిస్తుంది . ఈ పత్రం [5] EFFResNet-ViT ను పరిచయం చేస్తుంది , ఇది వైద్య చిత్ర వర్గీకరణ కోసం విజన్ ట్రాన్స్‌ఫార్మర్‌లతో కన్వల్యూషనల్ నెట్‌వర్క్‌లను కలిపే ఫ్యూజన్ మోడల్. ఇది పారదర్శకతను అందించడానికి మరియు క్లినికల్ నిర్ణయ మద్దతును మెరుగుపరచడానికి వివరించదగిన AI పద్ధతులను ప్రభావితం చేస్తుంది. మెదడు MRI లో మల్టీమోడల్ గ్లియోమా సెగ్మెంటేషన్ కోసం విజన్ ట్రాన్స్‌ఫార్మర్లు మరియు కన్వల్యూషనల్ నెట్‌వర్క్‌లను కలిపే హైబ్రిడ్ మోడల్‌ను ఈ పత్రం [6] అందిస్తుంది. ట్రాన్స్‌ఫార్మర్ ఆధారిత మెడికల్ ఇమేజింగ్ మోడల్‌లలో అటెన్షన్ మ్యాప్‌ల నుండి పొందిన దృశ్య వివరణల ప్రభావాన్ని ఈ పత్రం [7] అంచనా వేస్తుంది. మెదడు వ్యాధి గుర్తింపు కోసం బదిలీ అభ్యాసంతో కలిపి విజన్ ట్రాన్స్‌ఫార్మర్‌ల వాడకాన్ని ఈ పత్రం [8] అన్వేషిస్తుంది.

3. పద్ధతి

మెషిన్ లెర్నింగ్ టెక్నిక్‌లను ఉపయోగించి కస్టమర్ల కదలికను అంచనా వేయడం మరియు వాటి ప్రభావాన్ని పోల్చడం అనే లక్ష్యాలను సాధించడానికి, ఒక క్రమబద్ధమైన పద్ధతిని అనుసరిస్తారు. ఈ పద్ధతిలో అనేక కీలక దశలు ఉన్నాయి: డేటా సేకరణ, డేటా ప్రీప్రాసెసింగ్, ఫీచర్ ఎంపిక, మోడల్ శిక్షణ మరియు మూల్యాంకనం, హైపర్‌పారామీటర్ ట్యూనింగ్ మరియు తులనాత్మక విశ్లేషణ. ప్రిడిక్టివ్ మోడల్‌ల ఖచ్చితత్వం మరియు విశ్వసనీయతను నిర్ధారించడంలో ప్రతి దశ కీలకమైనది.

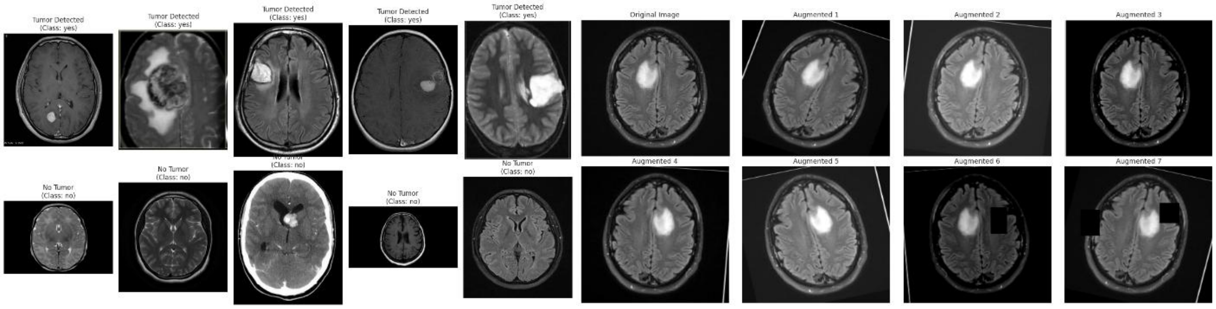
4. ప్రయోగాత్మక సెటప్ మరియు అమలు

PyTorch మరియు TensorFlow లైబ్రరీలను ఉపయోగించి Python లో ప్రయోగాత్మక సెటప్ అమలు చేయబడింది . అమలు ఒక నిర్మాణాత్మక వర్క్‌ఫ్లోను అనుసరించింది. మొదట, 253 చిత్రాలను కలిగి ఉన్న MRI డేటాసెట్ (155 "అవును" కణితిగా మరియు 98 "లేదు" కణితిగా లేబుల్ చేయబడింది) లోడ్

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

చేయబడింది మరియు ప్రీప్రాసెస్ చేయబడింది . సాధారణీకరణను మెరుగుపరచడానికి, రీసైజ్, క్షితిజసమాంతర ఫ్లిప్, రొటేషన్, యాదృచ్ఛిక ప్రకాశం/కాంట్రాస్ట్, ఎలాస్టిక్ ట్రాన్స్‌ఫార్మ్, కోర్స్ డ్రాప్అవుట్ మరియు నార్మలైజేషన్‌తో సహా అనేక డేటా ఆగ్మెంటేషన్ పద్ధతులు వర్తింపజేయబడ్డాయి. విజన్ ట్రాన్స్ ఫార్మర్ (ViT) మోడల్‌కు హైపర్‌పారామీటర్‌లతో శిక్షణ సెట్‌పై శిక్షణ ఇవ్వబడింది EPOCHS = 15, అభ్యాస రేటు = 1e-4, మరియు బరువు క్షీణత = 1e-6, 0.9412 యొక్క ఉత్తమ ధ్రువీకరణ ఖచ్చితత్వాన్ని సాధించింది. తరువాత మోడల్‌ను ప్రామాణిక పనితీరు మెట్రిక్‌లను ఉపయోగించి పరీక్ష సెట్‌లో మూల్యాంకనం చేశారు. చివరగా, నమూనా పరీక్ష చిత్రాల కోసం శ్రద్ధ పటాలు రూపొందించబడ్డాయి, ఇక్కడ ఫలితాలను అర్థం చేసుకోవడానికి వివరించదగిన AI పద్ధతులు వర్తింపజేయబడ్డాయి, మోడల్ యొక్క నిర్ణయం తీసుకునే ప్రక్రియపై అంతర్దృష్టులను అందిస్తాయి.



(a) (బి)

Fi g.1(a) డేటాసెట్ నుండి ఉదాహరణ MRI చిత్రాలు. (b) డేటా వృద్ధి తర్వాత MRI చిత్రాలు

5. ఫలితాల విశ్లేషణ

వర్గీకరణ పనిలో ViT మోడల్ బలమైన పనితీరును సాధించింది. గందరగోళం మ్యాట్రిక్స్ పట్టిక 1లో సంగ్రహించబడింది. CNN బేస్‌లైన్ మరియు Vi T + XAI నమూనాల మధ్య పనితీరు పోలిక వర్గీకరణ ఖచ్చితత్వంలో గణనీయమైన మెరుగుదలలను హైలైట్ చేస్తుంది. CNN బేస్‌లైన్ కోసం, మోడల్ 15 ట్రూ నెగటివ్‌లు (TN) మరియు 17 ట్రూ పాజిటివ్‌లు (TP) సాధించింది, కానీ ఇది 5 ఫాల్స్ పాజిటివ్‌లు (FP) మరియు 3 ఫాల్స్ నెగటివ్‌లు (FN) కూడా ఉత్పత్తి చేసింది, ఇది మితమైన తప్పుడు వర్గీకరణ రేట్లను సూచిస్తుంది. దీనికి విరుద్ధంగా, Vi T + XAI మోడల్ 18 TN మరియు 30 TPతో బలమైన పనితీరును చూపించింది, అయితే లోపాలను 2 FP మరియు 1 FN కి మాత్రమే తగ్గించింది. మొత్తంమీద, ViT + XAI విధానం CNN బేస్‌లైన్‌ను అధిగమించింది, అధిక సరైన గుర్తింపులను (TN మరియు TP రెండూ) మరియు తక్కువ తప్పుడు వర్గీకరణలను (FP మరియు FN) అందించింది, వర్గీకరణ పనులలో దాని ఉన్నతమైన విశ్వసనీయత మరియు వివరణాత్మకతను ప్రదర్శించింది.

ISSN: 2583-9055

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

పట్టిక 1: గందరగోళం మాతృక

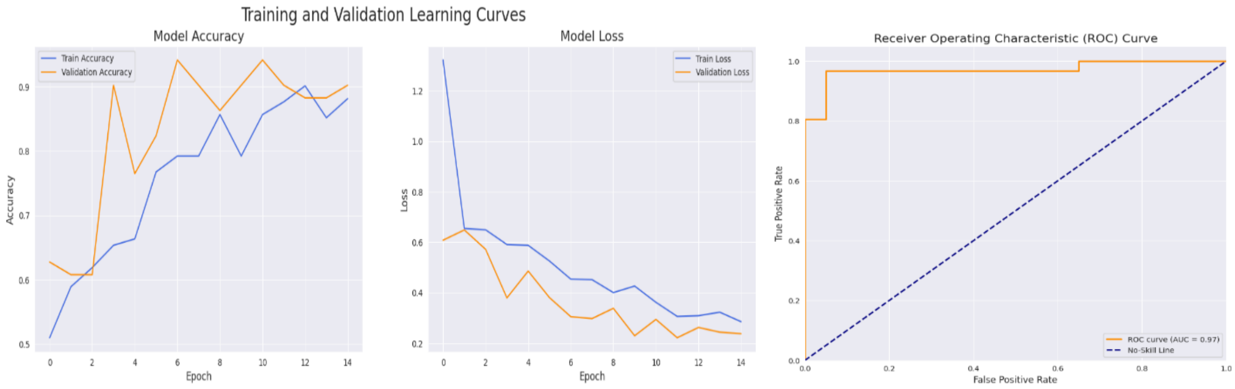
	నిజమైన ప్రతికూలత	తప్పుడు పాజిటివ్	తప్పుడు ప్రతికూలత	నిజమైన సానుకూలత
CNN బేస్‌లైన్	15	5	3	17
విఐటి + ఎక్స్ఐఐ	18	2	1. 1.	30 లు

చిత్రం 2(a) 15 యుగాలలో మోడల్ పనితీరు కోసం శిక్షణ మరియు ధ్రువీకరణ అభ్యాస వక్రతలను చూపిస్తుంది. చిత్రం 2(a) ఎడమ (మోడల్ ఖచ్చితత్వం): శిక్షణ ఖచ్చితత్వం యుగాలతో క్రమంగా పెరుగుతుంది, 0.6 చుట్టూ ప్రారంభమై చివరికి దాదాపు 0.9కి చేరుకుంటుంది. ధ్రువీకరణ ఖచ్చితత్వం కూడా మెరుగుపడుతుంది, యుగాలలో కొద్దిగా హెచ్చుతగ్గులకు లోనవుతుంది కానీ సాధారణంగా అదే పైకి ధోరణిని అనుసరిస్తుంది, చిన్న వైవిధ్యాలతో మంచి సాధారణీకరణను సూచిస్తుంది. చిత్రం 2(a) కుడి (మోడల్ నష్టం): శిక్షణ నష్టం 1.2 పైన నుండి 0.3 కంటే తక్కువకు స్థిరంగా తగ్గుతుంది, అయితే ధ్రువీకరణ నష్టం కూడా కొన్ని హెచ్చుతగ్గులతో తగ్గుతుంది, 0.4–0.5 చుట్టూ స్థిరీకరిస్తుంది. ఇది గణనీయమైన ఓవర్ ఫిట్టింగ్ లేకుండా ప్రభావవంతమైన ఆప్టిమైజేషన్‌ను సూచిస్తుంది. అభ్యాస వక్రతలు మోడల్ శిక్షణ మరియు ధ్రువీకరణ రెండింటికీ ప్రగతిశీల ఖచ్చితత్వ లాభాలను మరియు తగ్గుతున్న నష్టాన్ని సాధిస్తుందని సూచిస్తున్నాయి, బలమైన అభ్యాస ప్రవర్తన మరియు నమ్మకమైన పనితీరును చూపుతుంది.

చిత్రం 2(b) మోడల్ మూల్యాంకనం కోసం రిసీవర్ ఆపరేటింగ్ క్యారెక్టరిస్టిక్ (ROC) వక్రరేఖను ప్రదర్శిస్తుంది. నారింజ రంగు వక్రరేఖ మోడల్ యొక్క ROC పనితీరును సూచిస్తుంది, 0.971 యొక్క ఏరియా అండర్ ది కర్వ్ (AUC) స్కోర్‌ను సాధిస్తుంది, ఇది అద్భుతమైన వర్గీకరణ సామర్థ్యాన్ని సూచిస్తుంది. నీలిరంగు గీత రేఖ 0.5 AUCతో వికర్ణంగా ఉంచబడిన నో-స్కిల్ వర్గీకరణ (యాదృచ్ఛిక అంచనా)ని సూచిస్తుంది. మోడల్ యొక్క ROC వక్రరేఖ నో-స్కిల్ రేఖకు గణనీయంగా పైన ఉంటుంది, ముఖ్యంగా తప్పుడు సానుకూల రేట్లను (FPR) చాలా తక్కువగా ఉంచుతూ అధిక ట్రూ పాజిటివ్ రేట్లను (TPR) నిర్వహిస్తుంది. మోడల్ బలమైన వివక్షత శక్తిని మరియు తరగతుల మధ్య తేడాను గుర్తించడంలో అధిక విశ్వసనీయతను ప్రదర్శిస్తుంది, 0.971 యొక్క AUC విలువ ద్వారా ప్రతిబింబించే దాదాపు పరిపూర్ణ పనితీరుతో.

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025



(a) (బి)

చిత్రం 2: (ఎ) శిక్షణ vs. ధ్రువీకరణ ఖచ్చితత్వం/నష్ట వక్రతలు. (బి) మోడల్ వివక్షత సామర్థ్యాన్ని చూపించే ROC వక్రత.

తులనాత్మక ఫలితాలు పట్టిక 2లో ప్రదర్శించబడ్డాయి. ప్రతిపాదిత విజన్ ట్రాన్స్‌ఫార్మర్ మోడల్ పనితీరును బెంచ్‌మార్క్ చేయడానికి దీనిని అమలు చేయబడిన బేస్‌లైన్ CNN-ఆధారిత విధానంతో పోల్చారు. CNN బేస్‌లైన్ మోడల్ మొత్తం 80% ఖచ్చితత్వాన్ని సాధించింది, 'కాదు' తరగతికి 83% మరియు 'అవును' తరగతికి 77% ఖచ్చితత్వంతో. దీని రీకాల్ విలువలు 75% (కాదు) మరియు 85% (అవును), అయితే F1-స్కోర్లు 79% (కాదు) మరియు 81% (అవును). ఇది సమతుల్యమైన కానీ మితమైన పనితీరును చూపిస్తుంది, అయినప్పటికీ దీనికి వివరణ లేదు. ViT + XAI మోడల్ 94% ఖచ్చితత్వంతో బేస్‌లైన్‌ను గణనీయంగా అధిగమించింది . ఇది 95% ఖచ్చితత్వం (కాదు) మరియు 94% ఖచ్చితత్వం (అవును) సాధించింది, వాటితో పాటు 90% రీకాల్ (కాదు) మరియు 97% రీకాల్ (అవును). దీని F1-స్కోర్లు 92% (కాదు) మరియు 95% (అవును). ముఖ్యంగా, ఇది శ్రద్ధ మ్యాప్‌ల ద్వారా వివరణను అందిస్తుంది, క్లినికల్ నమ్మకాన్ని పెంచుతుంది. ViT + XAI మోడల్ అన్ని మెట్రిక్‌లలో వర్గీకరణ పనితీరును మెరుగుపరచడమే కాకుండా వివరణాత్మకతను (శ్రద్ధ పటాలు) కూడా అనుసంధానిస్తుంది, ఇది CNN బేస్‌లైన్‌తో పోలిస్తే వైద్య నిర్ణయ మద్దతుకు మరింత నమ్మదగినదిగా చేస్తుంది .

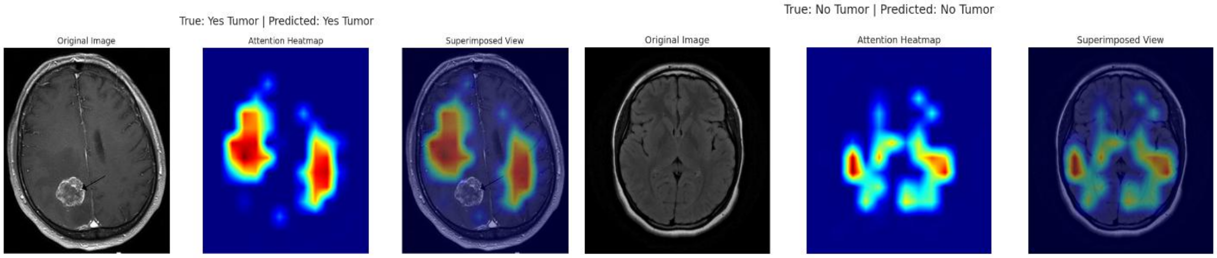
CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

పట్టిక 2: CNN బేస్‌లైన్ మరియు ప్రతిపాదిత ViT మోడల్ మధ్య తులనాత్మక పనితీరు.

మోడల్	అకు జాతి వివక్ష త	ప్రెసిషన్ లేదు	ప్రెసిషన్ అవును	రీకాల్ లేదు	రీకాల్ అవును	F1-స్కోరు లేదు	F1-స్కోరు అవును	అర్థవివర ణ
CNN బేస్‌లైన్	80%	83%	77%	75%	85%	79%	81%	ఉత్తర అమెరికా
విఐటి + ఎక్స్ఐఐ	94%	95%	94%	90%	97%	92%	95%	అటెన్షన్ మ్యాప్

అత్తి 3 లోని అటెన్షన్ మ్యాప్‌లు ViT మోడల్ కణితి ప్రాంతాలకు శ్రద్ధ వహిస్తుందని, దాని నిర్ణయాలు వైద్యపరంగా సంబంధిత ప్రాంతాలకు అనుగుణంగా ఉన్నాయని ధృవీకరిస్తాయని చూపిస్తుంది. అత్తి 3 (ఎడమ) MRI స్కాన్‌లలో మెదడు కణితి గుర్తింపు కోసం అటెన్షన్ విజువలైజేషన్‌తో వివరించదగిన AI వాడకాన్ని వివరిస్తుంది . ఎడమ చిత్రం అసలు MRI స్కాన్‌ను చూపిస్తుంది, ఇక్కడ కణితి ప్రాంతం కనిపిస్తుంది. మధ్య చిత్రం అటెన్షన్ హీట్‌మ్యాప్‌ను ప్రదర్శిస్తుంది, మోడల్ నిర్ణయం తీసుకోవడంలో అత్యంత ప్రభావవంతమైన ప్రాంతాలను హైలైట్ చేస్తుంది. ఎరుపు మరియు పసుపు ప్రాంతాలు అధిక అటెన్షన్‌కు అనుగుణంగా ఉంటాయి, కణితి ప్రాంతంపై సరిగ్గా దృష్టి పెడతాయి. కుడి చిత్రం అసలు MRI స్కాన్‌పై హీట్‌మ్యాప్‌ను అతివ్యాప్తి చేస్తుంది, మోడల్ యొక్క ఫోకస్‌ను వాస్తవ కణితి స్థానంతో స్పష్టంగా సమలేఖనం చేసే సూపర్‌పోజ్డ్ వ్యూను అందిస్తుంది. మోడల్ “అవును ట్యూమర్ ” (నిజమైన పాజిటివ్) ను సరిగ్గా అంచనా వేసింది మరియు అటెన్షన్ మ్యాప్ దానిని ప్రదర్శిస్తుంది.



చిత్రం 3: కణితి ప్రాంతాలను హైలైట్ చేసే అటెన్షన్ మ్యాప్ విజువలైజేషన్. (ఎడమ) అవును కణితి. (కుడి) కణితి లేదు

సరిగ్గా వర్గీకరించబడిన ప్రతికూల MRI కేసు (కణితి లేదు) కోసం వివరించదగిన AI విజువలైజేషన్‌ను అత్తి 3 (కుడి) ప్రదర్శిస్తుంది. ఎడమ చిత్రం కనిపించే కణితిని చూపించని అసలు MRI స్కాన్ . మధ్య చిత్రం అటెన్షన్ హీట్‌మ్యాప్‌ను ప్రదర్శిస్తుంది, ఇక్కడ హైలైట్ చేయబడిన ప్రాంతాలు (ఎరుపు/పసుపు

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025

రంగులో) అంచనా సమయంలో మోడల్ పరిగణించిన ప్రాంతాలను సూచిస్తాయి. ఈ ఫోకస్ పాయింట్లు చెల్లాచెదురుగా ఉన్నాయి కానీ కణితి లాంటి నిర్మాణాలకు అనుగుణంగా లేవు. కుడి చిత్రం MRI స్కాన్ లోని హీట్‌మ్యాప్‌ను అతివ్యాప్తి చేస్తుంది, మోడల్ సంబంధిత ప్రాంతాలను పరిశీలించినప్పటికీ కణితి లేకపోవడాన్ని సరిగ్గా నిర్ణయించిందని నిర్ధారిస్తుంది . మోడల్ " కణితి లేదు" (నిజమైన ప్రతికూలత) ను ఖచ్చితంగా అంచనా వేసింది మరియు అటెన్షన్ మ్యాప్ కణితి లేని ప్రాంతాలపై తప్పుగా దృష్టి పెట్టకుండా నిర్ణయం తీసుకున్నట్లు చూపించడం ద్వారా అర్థవివరణను బలోపేతం చేస్తుంది .

ముగింపు

ఈ అధ్యయనం విజన్ ట్రాన్స్‌ఫార్మర్లు మెదడు కణితి గుర్తింపుకు అధిక వర్గీకరణ ఖచ్చితత్వాన్ని మాత్రమే కాకుండా, శ్రద్ధ పటాల ద్వారా వివరణను కూడా అందిస్తాయని నిరూపిస్తుంది. ఈ మోడల్ క్లినికల్‌గా ముఖ్యమైన కణితి ప్రాంతాలపై దృష్టి పెడుతుందని నిర్ధారించడం ద్వారా విజువలైజేషన్‌లు ఆటోమేటెడ్ డయాగ్నస్టిక్ సిస్టమ్‌లపై నమ్మకాన్ని పెంచుతాయి. భవిష్యత్ పనిలో మల్టీ-సెంటర్ MRI స్కాన్‌లతో డేటాసెట్‌ను విస్తరించడం, హైబ్రిడ్ ViT -CNN ఆర్కిటెక్చర్‌లను అన్వేషించడం మరియు అధునాతన వివరణాత్మక పద్ధతులను సమగ్రపరచడం వంటివి ఉంటాయి. అదనంగా, ప్రతిపాదిత మోడల్ యొక్క ఆచరణాత్మక ప్రయోజనాన్ని అంచనా వేయడానికి క్లినికల్ వర్క్‌ఫ్లోలలో నిజ-సమయ విస్తరణను పరిశోధించవచ్చు.

ప్రస్తావనలు

1. Mzoughi, H., Njeh , I., BenSlima , M., Farhat, N., & Mhiri , C . (2025). MRI మెదడు ప్రాథమిక కణితుల చిత్రాల కోసం విజన్ ట్రాన్స్‌ఫార్మర్లు (Vi T) మరియు డీప్ కన్వల్యూషనల్ న్యూరల్ నెట్‌వర్క్ (D-CNN) ఆధారిత నమూనాలు వివరించదగిన కృత్రిమ మేధస్సు (XAI) ద్వారా మద్దతు ఇవ్వబడిన బహుళ-వర్గీకరణ. *ది విజువల్ కంప్యూటర్* , *41* (4), 2123-2142.
2. హోస్నీ, కె.ఎం., & మొహమ్మద్, ఎం.ఎ. (2025). మెదడు కణితిని గుర్తించడం మరియు వర్గీకరించడం కోసం వివరించదగిన AI మరియు దృష్టి ట్రాన్స్‌ఫార్మర్లు: ఒక సమగ్ర సర్వే. *ఆర్టిఫిషియల్ ఇంటెలిజెన్స్ రివ్యూ* , *58* (9), 259.
3. అయ్యు, ఎజె, వాని, ఎన్., రమణి, ఎం., కుమార్, ఎ., పంత్, ఎస్., కోటేచా, కె., ... & అల్- దానఖ్ , ఎ. (2025). మెదడు కణితి గుర్తింపు కోసం ఆప్టిమైజ్ చేయబడిన లోతైన అభ్యాసం: శ్రద్ధ విధానాలు మరియు క్లినికల్ వివరణతో కూడిన హైబ్రిడ్ విధానం. *సైంటిఫిక్ రిపోర్ట్స్* , *15* (1), 31386.
4. రాజ్‌పూత్, ఆర్., జైన్, ఎస్., & సెమ్వాల్, విబి (2025). బయోట్రాన్స్‌ఎక్స్ : వివరించదగిన అంతర్దృష్టులతో మెదడు కణితి వర్గీకరణ కోసం ద్వి-స్థాయి రూటింగ్ శ్రద్ధతో కూడిన నవల ద్వి-పూర్వ ఆధారిత హైబ్రిడ్ మోడల్. *కంప్యూటర్స్ ఇన్ బయాలజీ అండ్ మెడిసిన్* , *195* , 110515.

CC BY

5. హుస్సేన్, టి., షోనో , హెచ్., హుస్సేన్, ఎ., హుస్సేన్, డి., ఇస్మాయిల్, ఎం., మీర్, టిహెచ్, ... & అఖీ , ఎస్ఎ (2025). EFFResNet-ViT : వివరించదగిన వైద్య చిత్ర వర్గీకరణ కోసం ఫ్యూజన్-ఆధారిత కన్వల్యూషనల్ మరియు విజన్ ట్రాన్స్‌ఫార్మర్ మోడల్. *IEEE యాక్సెస్* .
6. జీనెల్డిన్ , RA , కరార్, ME , ఎల్షేర్, Z ., కోబర్గర్ , J ., విర్ట్జ్, CR , బర్గర్ట్, O ., & మాథిస్-ఉల్రిచ్, F . (2024). మెదడు MRI లో మల్టీమోడల్ గ్లియోమా సెగ్మెంటేషన్ కోసం వివరించదగిన హైబ్రిడ్ విజన్ ట్రాన్స్‌ఫార్మర్లు మరియు కన్వల్యూషనల్ నెట్‌వర్క్. *సైంటిఫిక్ రిపోర్ట్స్* , *14* (1), 3713.
7. చుంగ్, ఎం., వోన్, జెబి, కిమ్, జి., కిమ్, వై., & ఓజ్బులాక్ , యు. (2024, అక్టోబర్). ట్రాన్స్‌ఫార్మర్-బేస్డ్ మెడికల్ ఇమేజింగ్ కోసం అటెన్షన్ మ్యాప్‌ల దృశ్య వివరణలను మూల్యాంకనం చేయడం. *మెడికల్ ఇమేజ్ కంప్యూటింగ్ మరియు కంప్యూటర్-అసిస్టెడ్ ఇంటర్వెన్షన్‌పై అంతర్జాతీయ సమావేశంలో* (పేజీలు 110-120). చామ్: స్ప్రింగర్ నేచర్ స్విట్జర్లాండ్.
8. సర్కర్, ఎస్., రెఫాట్, ఎస్ఆర్, ప్రియోటీ , ఎఫ్ఎఫ్, ఇస్లాం, ఎస్., ముహమ్మద్, టి., & హోక్, ఎంఎ (2024, డిసెంబర్). మెదడు వ్యాధి గుర్తింపు కోసం విజన్ ట్రాన్స్‌ఫార్మర్ మరియు ట్రాన్స్‌ఫర్ లెర్నింగ్‌ను పరిశోధించడం మరియు వివరించడం కోసం ఒక అన్వేషణాత్మక విధానం. *2024లో కంప్యూటర్ మరియు ఇన్ఫర్మేషన్ టెక్నాలజీపై 27వ అంతర్జాతీయ సమావేశం (ICCIT)* (పేజీలు 3224-3229). IEEE.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025
PP 72-78

Interpretable Brain Tumor Classification using Vision Transformers with Attention Maps

Dr Humera Shaziya
Email-humera.shaziya@gmail.com
Department of Informatics, Nizam College, Osmania University

Abstract

This paper presents an interpretable deep learning approach for brain tumor classification using Vision Transformers (ViT). The study emphasizes not only classification performance but also explainability by visualizing attention maps. A brain MRI dataset with imbalanced classes was used, and data augmentation was applied to improve generalization. The proposed ViT-based model was trained and evaluated using standard metrics, including Accuracy, Precision, Recall, F1-Score, and AUC/ROC. Furthermore, attention map visualization demonstrates how the model focuses on tumor regions, enhancing trustworthiness and clinical usability. Experimental results indicate that the ViT architecture achieves high classification performance with an accuracy of 94% while providing interpretable outputs that support decision-making in medical imaging whereas the CNN baseline model achieves 80% accuracy.

Keywords

Brain Tumor Classification, Vision Transformer, Explainable AI, Attention Maps, Medical Imaging, MRI

Introduction

Brain tumors are life-threatening conditions requiring timely diagnosis and effective treatment. Medical imaging, especially MRI scans, plays a crucial role in early detection. Traditional machine learning methods rely heavily on handcrafted features, limiting their ability to generalize across diverse tumor patterns. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved classification accuracy. However, CNNs often lack interpretability, a critical requirement in clinical applications. Vision Transformers (ViTs) have emerged as powerful alternatives to CNNs, leveraging self-attention mechanisms to capture global dependencies in images. Importantly, the attention maps generated by ViTs provide a means for explainable AI (XAI) by showing which image regions influenced the model's decision. In this study, we propose an interpretable brain tumor classification model using ViTs and visualize its attention maps to evaluate the trustworthiness of predictions.

CC BY

Research Objectives and Methodology

The proposed methodology adopts a structured approach for brain tumor classification using MRI scans. The dataset comprises brain MRI images encompassing multiple tumor classes, with an inherent imbalance in class distribution. To address this, data preprocessing was performed. The core model architecture employed is a fine-tuned Vision Transformer (ViT). Model performance was evaluated to ensure a comprehensive understanding of predictive capability. Furthermore, interpretability was integrated into the framework by extracting attention maps from the trained ViT, thereby enabling visualization of the model's focus regions and offering explainable insights into the decision-making process.

2. Literature Survey

The paper [1] proposes hybrid Vision Transformer (ViT) and Deep CNN-based models for multi-class classification of MRI brain tumor images. It also integrates Explainable AI (XAI) techniques to enhance model transparency and clinical trust. Paper [2] presents a comprehensive survey on the use of Vision Transformers (ViTs) and Explainable AI (XAI) in brain tumor detection and classification. The paper [3] introduces an optimized hybrid deep learning approach incorporating attention mechanisms for accurate brain tumor detection. Paper [4] proposes BioTransX, a bi-former-based hybrid model employing bi-level routing attention for brain tumor classification. The paper [5] introduces EFFResNet-ViT, a fusion model combining convolutional networks with Vision Transformers for medical image classification. It leverages explainable AI techniques to provide transparency and improve clinical decision support. The paper [6] presents a hybrid model combining Vision Transformers and convolutional networks for multimodal glioma segmentation in brain MRI. The paper [7] evaluates the effectiveness of visual explanations derived from attention maps in transformer-based medical imaging models. The paper [8] explores the use of Vision Transformers combined with transfer learning for brain disease detection.

3. Methodology

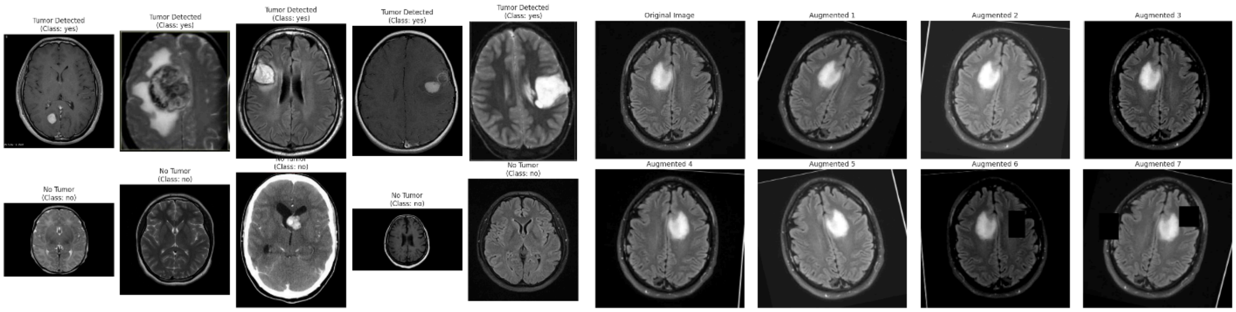
To achieve the objectives of predicting customer churn using machine learning techniques and comparing their effectiveness, a systematic methodology is followed. This methodology includes several key steps: data collection, data preprocessing, feature selection, model training and evaluation, hyperparameter tuning, and comparative analysis. Each step is crucial in ensuring the accuracy and reliability of the predictive models.

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025
PP 72-78

4. Experimental Setup and Implementation

The experimental setup was implemented in Python using PyTorch and TensorFlow libraries. The implementation followed a structured workflow. First, the MRI dataset containing 253 images (with 155 labeled as "Yes" tumor and 98 labeled as "No" tumor) was loaded and preprocessed. To improve generalization, several data augmentation techniques were applied, including Resize, Horizontal Flip, Rotation, Random Brightness/Contrast, Elastic Transform, Coarse Dropout, and Normalization. The Vision Transformer (ViT) model was then trained on the training set with hyperparameters EPOCHS = 15, learning rate = 1e-4, and weight decay = 1e-6, achieving a best validation accuracy of 0.9412. The model was subsequently evaluated on the test set using standard performance metrics. Finally, attention maps were generated for sample test images, where Explainable AI techniques were applied to interpret the results, providing insights into the model's decision-making process.



(a) (b)

Fig.1(a) Example MRI images from the dataset. (b) MRI images post data augmentation

5. Result Analysis

The ViT model achieved strong performance on the classification task. Confusion Matrix is summarized in Table 1. The performance comparison between the CNN Baseline and ViT + XAI models highlights significant improvements in classification accuracy. For the CNN Baseline, the model achieved 15 True Negatives (TN) and 17 True Positives (TP), but it also produced 5 False Positives (FP) and 3 False Negatives (FN), indicating moderate misclassification rates. In contrast, the ViT + XAI model showed stronger performance with 18 TN and 30 TP, while reducing errors to only 2 FP and 1 FN. Overall, the ViT + XAI approach outperformed the CNN Baseline, offering higher correct detections (both TN and TP) and fewer misclassifications (FP and FN), demonstrating its superior reliability and explainability in classification tasks.

CC BY

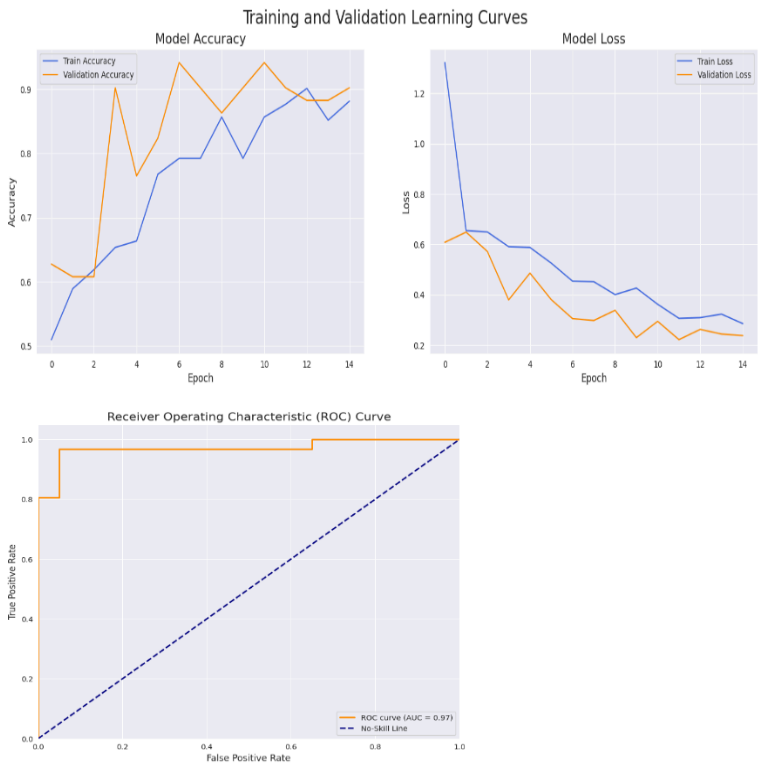
Table 1: Confusion Matrix

	True Negative	False Positive	False Negative	True Positive
CNN Baseline	15	5	3	17
ViT + XAI	18	2	1	30

The fig 2(a) shows the training and validation learning curves for model performance across 15 epochs. Fig 2(a) left (Model Accuracy): Training accuracy steadily increases with epochs, starting around 0.6 and reaching nearly 0.9 by the end. Validation accuracy also improves, fluctuating slightly across epochs but generally follows the same upward trend, indicating good generalization with minor variations. Fig 2(a) right (Model Loss): Training loss decreases consistently from above 1.2 to below 0.3, while validation loss also declines with some fluctuations, stabilizing around 0.4–0.5. This suggests effective optimization without significant overfitting. The learning curves indicate that the model achieves progressive accuracy gains and decreasing loss for both training and validation, showing strong learning behavior and reliable performance.

The fig 2(b) displays the Receiver Operating Characteristic (ROC) curve for model evaluation. The orange curve represents the ROC performance of the model, achieving an Area Under the Curve (AUC) score of 0.971, which indicates excellent classification capability. The blue dashed line represents the No-Skill classifier (random guess), positioned diagonally with an AUC of 0.5. The ROC curve of the model lies significantly above the No-Skill line, particularly maintaining high True Positive Rates (TPR) while keeping False Positive Rates (FPR) very low. The model demonstrates strong discriminative power and high reliability in distinguishing between classes, with near-perfect performance as reflected by the AUC value of 0.971.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025
PP 72-78



(a) (b)

Fig. 2: (a) Training vs. validation accuracy/loss curves. (b) ROC curve showing model discrimination ability.

Comparative results are presented in table 2. To benchmark the performance of the proposed Vision Transformer model has been compared it with a baseline CNN-based approach implemented. The CNN Baseline model achieved an overall accuracy of 80%, with a precision of 83% for the ‘No’ class and 77% for the ‘Yes’ class. Its recall values were 75% (No) and 85% (Yes), while the F1-scores were 79% (No) and 81% (Yes). This shows balanced but moderate performance, though it lacks interpretability. The ViT + XAI model significantly outperformed the baseline with an accuracy of 94%. It attained 95% precision (No) and 94% precision (Yes), along with 90% recall (No) and 97% recall (Yes). Its F1-scores were 92% (No) and 95% (Yes). Importantly, it provides interpretability through attention maps, enhancing clinical trust. The ViT + XAI model not only improves classification performance across all metrics but also integrates explainability (attention maps), making it more reliable for medical decision support compared to the CNN Baseline.

Table 2: Comparative performance between CNN baseline and proposed ViT model.

BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025
PP 72-78

Model	Accuracy	Precision No	Precision Yes	Recall No	Recall Yes	F1-Score No	F1-Score Yes	Interpretability
CNN Baseline	80%	83%	77%	75%	85%	79%	81%	NA
ViT + XAI	94%	95%	94%	90%	97%	92%	95%	Attention Map

The attention maps in fig 3 show that the ViT model attends to the tumor regions, validating that its decisions align with medically relevant areas. The fig3 (left) illustrates the use of explainable AI with attention visualization for brain tumor detection in MRI scans. The left image shows the original MRI scan, where the tumor region is visible. The middle image presents the attention heatmap, highlighting the regions most influential in the model's decision-making. The red and yellow areas correspond to high attention, correctly focusing on the tumor region. The right image overlays the heatmap on the original MRI scan, providing a superimposed view that clearly aligns the model's focus with the actual tumor location. The model correctly predicted "Yes Tumor" (true positive) and the attention map demonstrates it.

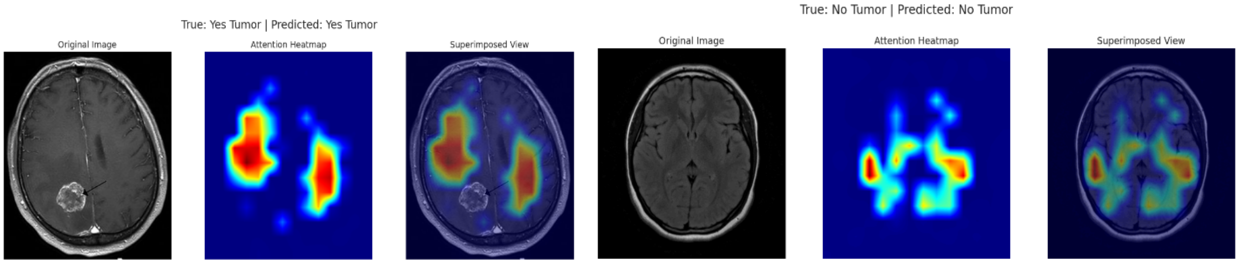


Fig. 3: Attention map visualization highlighting tumor regions.(left)yes tumor.(Right)No Tumor

The fig 3 (right) demonstrates explainable AI visualization for a correctly classified negative MRI case (no tumor). The left image is the original MRI scan showing no visible tumor. The middle image displays the attention heatmap, where highlighted regions (in red/yellow) indicate areas considered by the model during prediction. These focus points are scattered but do not correspond to tumor-like structures. The right image overlays the heatmap on the MRI scan, confirming that the model examined relevant regions yet correctly determined the absence of a tumor. The model accurately predicted "No Tumor" (true negative), and the attention map reinforces interpretability by showing that the decision was made without incorrectly focusing on non-tumor regions.

Conclusion

CC BY

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, Nov 2025
PP 72-78

This study demonstrates that Vision Transformers provide not only high classification accuracy for brain tumor detection but also interpretability through attention maps. The visualizations enhance trust in automated diagnostic systems by confirming that the model focuses on clinically significant tumor regions. Future work includes expanding the dataset with multi-center MRI scans, exploring hybrid ViT-CNN architectures, and integrating advanced interpretability techniques. Additionally, real-time deployment in clinical workflows can be investigated to evaluate the practical utility of the proposed model.

References

1. Mzoughi, H., Njeh, I., BenSlima, M., Farhat, N., & Mhiri, C. (2025). Vision transformers (ViT) and deep convolutional neural network (D-CNN)-based models for MRI brain primary tumors images multi-classification supported by explainable artificial intelligence (XAI). *The Visual Computer*, *41*(4), 2123-2142.
2. Hosny, K. M., & Mohammed, M. A. (2025). Explainable AI and vision transformers for detection and classification of brain tumor: a comprehensive survey. *Artificial Intelligence Review*, *58*(9), 259.
3. Aiya, A. J., Wani, N., Ramani, M., Kumar, A., Pant, S., Kotecha, K., ... & Al-Danakh, A. (2025). Optimized deep learning for brain tumor detection: a hybrid approach with attention mechanisms and clinical explainability. *Scientific Reports*, *15*(1), 31386.
4. Rajpoot, R., Jain, S., & Semwal, V. B. (2025). BioTransX: A novel bi-former based hybrid model with bi-level routing attention for brain tumor classification with explainable insights. *Computers in Biology and Medicine*, *195*, 110515.
5. Hussain, T., Shouno, H., Hussain, A., Hussain, D., Ismail, M., Mir, T. H., ... & Akhy, S. A. (2025). EFFResNet-ViT: A fusion-based convolutional and vision transformer model for explainable medical image classification. *IEEE Access*.
6. Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2024). Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI. *Scientific reports*, *14*(1), 3713.
7. Chung, M., Won, J. B., Kim, G., Kim, Y., & Ozbulak, U. (2024, October). Evaluating Visual Explanations of Attention Maps for Transformer-Based Medical Imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 110-120). Cham: Springer Nature Switzerland.
8. Sarker, S., Refat, S. R., Preotee, F. F., Islam, S., Muhammad, T., & Hoque, M. A. (2024, December). An Exploratory Approach Towards Investigating and Explaining Vision Transformer and Transfer Learning for Brain Disease Detection. In *2024 27th International Conference on Computer and Information Technology (ICCIT)* (pp. 3224-3229). IEEE.

CC BY