Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated classification of tau neuropathology from histological images using vision-based machine learning

Lily Sharman
UCL Queen Square Institute of Neurology
Andy Choi
UCL Queen Square Institute of Neurology
Sriya Sreekantam
UCL Queen Square Institute of Neurology
Charvi Varampati
UCL Queen Square Institute of Neurology
Gabriela Caballero
UCL Queen Square Institute of Neurology
Toby Curless
UCL Queen Square Institute of Neurology
Zannatul Farin
UCL Queen Square Institute of Neurology
Mariam Bechtawi
UCL Queen Square Institute of Neurology
Nancy Chiraki
UCL Queen Square Institute of Neurology
Hemanth R. Nelvagal
UCL Queen Square Institute of Neurology
Yau Mun Lim
UCL Queen Square Institute of Neurology
Sebastian Brandner
UCL Queen Square Institute of Neurology
Thomas T. Warner
UCL Queen Square Institute of Neurology
Zane Jaunmuktane
UCL Queen Square Institute of Neurology
Patrick W. Cullinane
patrick.cullinane.15@ucl.ac.uk

UCL Queen Square Institute of Neurology

Article

Keywords:

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11028200/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Primary tauopathies are neurodegenerative diseases characterised by accumulation of misfolded, hyperphosphorylated tau protein in neuronal and glial cells, each defined by hallmark inclusion morphologies. Post-mortem neuropathological examination remains the diagnostic gold standard but relies on scarce specialist expertise, and existing computational approaches are more developed for quantifying overall tau pathology burden in histological sections than for resolving cellular context. We evaluated whether Google Teachable Machine, a freely available no-code computer vision platform, could classify tauopathies and their associated tau inclusions. Two classifiers were trained on phosphorylated tau-immunostained sections from 125 cases: one classifying pancortical pathology patterns across eight distinct tauopathies at 5× magnification, the other classifying nine individual tau inclusion morphologies at 40× magnification. Case-level accuracy was 73.5% and 81.8% for the low- and high-magnification models respectively, with image-level accuracies of 77.4% and 72.2%. Recall reached 90% or higher for Alzheimer's disease, neurofibrillary tangles and granular/fuzzy astrocytes, with favourable area under the curve values across most categories. Performance was reduced in rare, underrepresented and morphologically overlapping categories, and where class boundaries depend on amyloid-β burden not visible on tau immunohistochemistry. These findings demonstrate that general-purpose vision models can be used for pattern recognition and morphology-based classification in neuropathology, providing an accessible benchmark for future approaches. Trained classifiers are provided as a web application.

Introduction

Tauopathies are a group of neurodegenerative diseases characterised by the abnormal accumulation of misfolded tau protein within neurones and glial cells[1]. Neuropathological diagnosis relies on the anatomical and cellular distribution and morphology of tau inclusions on immunostained tissue sections[2]. Biochemically, they can be classified according to the predominant isoforms of tau present in filaments, namely 3-repeat (3R), 4-repeat (4R), or mixed 3R/4R tau, as well as the molecular structure of filament cores[3]. The principal 3R tauopathy is Pick's disease (PiD), characterised by neuronal and glial Pick bodies and ramified astrocytes (RA)[1]. The major 4R tauopathies include progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and globular glial tauopathy (GGT), characterised respectively by tufted astrocytes (TA), astrocytic plaques (AP) and globular glial inclusions[4] as well as ageing-related tau astrogliopathy (ARTAG), characterised by predominantly limbic granular/fuzzy astrocyte (GFA) and thorn-shaped astrocyte (TSA) pathology. Mixed 3R/4R tauopathies include Alzheimer's disease (AD) and primary age-related tauopathy (PART), characterised by neurofibrillary tangles (NFTs), and chronic traumatic encephalopathy (CTE), characterised by distinctive cortical subpial and perivascular TSA tau pathology[5,6]. Advances in artificial intelligence provide opportunities to automate aspects of visual assessment, supporting specialist neuropathological evaluation and facilitating reproducible analysis at scale[7]. Computational approaches quantifying total phosphorylated tau (pTau) immunostaining provide reproducible measures of pathology burden but limited information

on morphology of cellular tau inclusions[8,9]. Quantification of total pTau burden is also insensitive to co-pathologies, particularly when distinct cellular inclusions are present within the same region of interests[10]. Morphology-based classification could address these limitations by distinguishing specific neuronal and glial inclusion types and may potentially identify subtle morphological variation associated with clinical phenotype, genotype or underlying tau strain.

General-purpose computer vision tools may provide an accessible means of applying morphology-based classification to tau pathology. Google Teachable Machine (GTM; Google LLC) uses transfer learning from pretrained convolutional neural networks (CNNs), allowing domain-specific image classification using relatively small datasets and without specialised computational infrastructure[11]. The underlying network, MobileNet, was pretrained on ImageNet ILSVRC-2012, a database of over a million natural images spanning 1,000 object categories; its convolutional layers are retained as a fixed feature extractor, and only a final classification layer is trained on the user's own images[12,13]. Previous studies have shown that GTM and comparable code-free platforms can extract clinically relevant information from histopathological images, including survival-associated features in melanoma, multiclass patterns in breast cancer, and discrimination of neoplastic from reactive glial lesions in central nervous system tissue[14–16]. However, its ability to distinguish neurodegenerative disease-associated tau pathology patterns and specific inclusion morphologies in post-mortem brain tissue has not been established. We therefore evaluated whether GTM could classify tau neuropathology at two spatial scales using pTau-immunostained tissue sections: a low-magnification model to distinguish disease-associated patterns in the frontal cortex, and a high-magnification model to recognise characteristic neuronal and astrocytic inclusion morphologies. We also developed a web-based application through which users can apply the trained classifier to new pTau-immunostained histology images. Together, these classifiers could provide an accessible and reproducible complement to specialist neuropathological assessment, support digital pathology pipelines for automated classification of tiles generated from histological sections and serve as a reference baseline for future task-specific and pathology foundation models designed to classify neuropathological patterns and inclusion morphologies across neurodegenerative tauopathies.

Results

Neuropathological cohort

The study included 125 cases comprising PiD (n = 19); PSP (n = 20); CBD (n = 18); GGT (n = 5); PART (n = 6); low level Alzheimer’s disease neuropathologic change (ADNC) (n = 7); intermediate level ADNC (n = 22); high level ADNC (n = 20); CTE (n = 3); Lewy body disease (LBD) (n = 3); frontotemporal lobar degeneration (FTLD) TDP-43 (n = 1); and Huntington’s disease (n = 1). The FTLD-TDP-43, Huntington’s disease and LBD cases were included only as sources of limbic GFAs, with LBD also providing limbic TSA tau pathology for training and testing the high-magnification model. The anatomical regions and tau inclusion types for each diagnostic group are summarised in Table 1.

Table 1
Neuropathological diagnoses, case numbers, anatomical regions, and tau morphological features represented in the study cohort.

Diagnosis	5x model	Extracted features	n	Anatomical region
PiD	Yes	RA	19	Frontal cortex, temporal cortex
PSP	Yes	TA, NFT	20	Frontal cortex, temporal cortex, parietal cortex
CBD	Yes	AP, NFT	18	Frontal cortex, temporal cortex, parietal cortex
GGT	Yes	GAI, NFT	5	Frontal cortex, temporal cortex, hippocampus
CTE	Yes	TSA, NFT	3	Frontal cortex, temporal cortex, hippocampus
Control, PART	Yes	TSA, GFA	6	Frontal cortex, temporal cortex, parietal cortex, hippocampus, amygdala
Control, low level ADNC	Yes	TSA, GFA	7	Frontal cortex, temporal cortex, parietal cortex, hippocampus, amygdala
Control, intermediate level ADNC	Yes	GFA, Ageing NP, AD NP, NFT	22	Frontal cortex, temporal cortex, parietal cortex, amygdala
High level ADNC	Yes	TSA, AD NP, NFT	20	Frontal cortex, temporal cortex, parietal cortex
Lewy body disease	No	TSA, GFA	3	Frontal cortex, temporal cortex, hippocampus, amygdala
FTLD-TDP-43	No	GFA	1	Amygdala
Huntington's disease	No	GFA	1	Amygdala

Cases marked "Yes" were included in both the low- and high-magnification models; those marked "No" were used only for the high-magnification model. Extracted features represent the morphological features sampled from each diagnostic group for training and testing of the high-magnification model. PiD, Pick's disease; PSP, progressive supranuclear palsy; CBD, corticobasal degeneration; GGT, globular glial tauopathy; CTE, chronic traumatic encephalopathy; PART, primary age-related tauopathy; ADNC, Alzheimer's disease neuropathological change, AD, FTLD-TDP-43, frontotemporal lobar degeneration with TDP-43 pathology; RA, ramified astrocyte; TA, tufted astrocyte; NFT, neurofibrillary tangle; AP, astrocytic plaque; GAI, globular astrocytic inclusion; TSA, thorn-shaped astrocyte; GFA, granular/fuzzy astrocyte; NP, neuritic plaque; AD, Alzheimer's disease.

Model training and convergence

Two GTM models were evaluated: a low-magnification model for classifying disease-associated pTau pathology patterns and a high-magnification model for classifying individual cell inclusion types. The composition of the training (including internal validation) and testing datasets for the low- and high-magnification models, including the numbers of cases and images represented in each category, is shown in Fig. 1. For the purposes of model classification, the low and intermediate ADNC groups were labelled 'Low AD', while the high ADNC was labelled 'AD'. During training, the low-magnification model reached a training accuracy of 99.5% against an internal validation accuracy of 85.6% by the final epoch, a generalisation gap of approximately 14 percentage points. Training loss decreased to 0.055 by the final epoch, whereas validation loss reached a minimum of 0.390 at epoch 28 before rising to 0.503 by epoch 49, indicating mild overfitting by this point. In contrast, the high-magnification model reached a final training accuracy of 94.1% against an internal validation accuracy of approximately 88.0%, resulting in a smaller generalisation gap of approximately 6 percentage points. Training loss decreased steadily to 0.025 by epoch 25 while validation loss remained broadly steady around 0.10 after declining rapidly within the first 6 epochs, suggesting less overfitting than in the low-magnification model (Supplementary Fig. 1).

GTM identified disease-associated tau patterns at low and high magnification

Overall image-level accuracy was higher for the low-magnification model than for the high-magnification model, at 77.4% and 72.2% respectively. Case-level predictions were derived by assigning each case to the disease group or morphological class predicted for the majority of its constituent images. On this basis the ordering was reversed, with accuracies of 81.8% for the high-magnification model and 73.5% for the low-magnification model.

At low magnification, all PART images were classified as Low AD, consistent with the shared pTau immunohistochemical pattern of these groups. When PART and Low AD were considered as a combined group, PART showed the highest recall, at 100%, followed by AD at 97.9%, with Low AD, CBD and PSP falling between 76.0% and 83.3% (Fig. 2a, Table 2). Recall was lower for GGT, PiD and CTE, at 52.6%, 36.7% and 12.0%, respectively. GGT was most frequently misclassified as PSP (47.4%), while CTE predictions were distributed predominantly across PSP (44.0%), CBD (28.0%) and GGT (16.0%). At low magnification, PiD was most frequently misclassified as AD.

Table 2
Class-specific performance of the low- and high-magnification tau pathology classification models.

Model	Class	Precision	Recall	F1 score	AUC
Low Magnification	PiD (n = 60)	0.917	0.367	0.534	0.948
	PSP (n = 150)	0.797	0.760	0.778	0.956
	CBD (n = 140)	0.882	0.807	0.843	0.972
	GGT (n = 19)	0.294	0.526	0.377	0.872
	CTE (n = 25)	0.600	0.120	0.200	0.894
	PART (n = 30)	1.000	1.000	1.000	0.903
	Low AD (n = 114)	0.720	0.833	0.772	0.931
	AD (n = 144)	0.762	0.979	0.857	0.989
High Magnification	RA (n = 400)	0.726	0.405	0.520	0.770
	TA (n = 400)	0.706	0.798	0.749	0.956
	AP (n = 400)	0.813	0.795	0.804	0.953

Precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC) are reported for each disease group or morphological class. The *n* under each class represents the number of test images evaluated for each class. PiD, Pick's disease; PSP, progressive supranuclear palsy; CBD, corticobasal degeneration; GGT, globular glial tauopathy; CTE, chronic traumatic encephalopathy; PART, primary age-related tauopathy; Low AD, low level Alzheimer's disease neuropathological change; AD, high level Alzheimer's disease neuropathological change; RA, ramified astrocyte; TA, tufted astrocyte; AP, astrocytic plaque; GAI, globular astrocytic inclusion; TSA, thorn-shaped astrocyte; GFA, granular/fuzzy astrocyte; NP, neuritic plaque; AD, Alzheimer's disease; NFT, neurofibrillary tangle.

Model	Class	Precision	Recall	F1 score	AUC
	GAI (n = 400)	0.535	0.210	0.302	0.601
	TSA (n = 400)	0.580	0.825	0.681	0.938
	GFA (n = 400)	0.782	0.950	0.858	0.989
	Ageing NP (n = 400)	0.673	0.798	0.730	0.925
	AD NP (n = 400)	0.832	0.818	0.825	0.931
	NFT (n = 400)	0.791	0.900	0.842	0.977

Precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC) are reported for each disease group or morphological class. The *n* under each class represents the number of test images evaluated for each class. PiD, Pick's disease; PSP, progressive supranuclear palsy; CBD, corticobasal degeneration; GGT, globular glial tauopathy; CTE, chronic traumatic encephalopathy; PART, primary age-related tauopathy; Low AD, low level Alzheimer's disease neuropathological change; AD, high level Alzheimer's disease neuropathological change; RA, ramified astrocyte; TA, tufted astrocyte; AP, astrocytic plaque; GAI, globular astrocytic inclusion; TSA, thorn-shaped astrocyte; GFA, granular/fuzzy astrocyte; NP, neuritic plaque; AD, Alzheimer's disease; NFT, neurofibrillary tangle.

At high magnification, recall was highest for GFA at 95.0%, followed by NFT at 90.0%, with TSA, high level Alzheimer's disease-associated neuritic plaque (AD NP), PSP-associated TA and CBD-associated AP pathology falling between 79.5% and 82.5% (Table 2). Low level ADNC-associated neuritic plaque (ageing NP) images were frequently classified as AD NP, reflecting that these two classes represent the same morphology on immunohistochemistry. When ageing NP and AD NP were considered as a combined class, ageing NP recall increased from 10.5% to 79.8%. Recall was lower for RAs in PiD and globular astrocytic inclusions (GAI) in GGT, at 40.5% and 21.0%, respectively. RA predictions were distributed across several classes, particularly TSAs (18.5%) and TAs (15.8%), whereas GAIs were most frequently misclassified as TSAs (36.8%) and NFTs (17.8%). The principal misclassification patterns are illustrated further in Fig. 3 and Fig. 4. Among the most misclassified images, representative low-magnification examples showed substantial visual similarity between CTE and GGT and the predicted PSP and CBD groups, particularly in the distribution and density of tau-positive pathology (Fig. 3). At high magnification, misclassified PiD-associated RA and GGT-associated GAI images frequently showed similar morphological features, particularly in cell shape and processes (Fig. 4).

Favourable precision with specific exceptions

Threshold-dependent precision and recall profiles are shown in Fig. 2b, with values at the default confidence threshold of 0.5 reported in Table 2. At low magnification, PiD showed the highest precision (91.7%), but low recall, indicating that positive predictions were generally reliable although many true PiD images were assigned to other groups. GGT showed the lowest precision (29.4%). Low level AD, AD, PSP and CBD showed intermediate precision values, ranging from 72.0% to 88.2%, with relatively balanced precision–recall profiles. At high magnification, precision was highest for AD NP (83.2%) and lowest for GGT-associated GAI (53.5%), which together with its low recall gave GAI the weakest overall balance of any inclusion class. TSA showed a similar imbalance in the opposite direction, combining high recall with a precision of 58.0%, whereas RA showed moderate precision at 72.6% against low recall. The remaining classes — AP, NFT, GFA and TA — fell between 70.6% to 81.3%; of these, AP and TA, along with AD NP, showed the most balanced precision–recall profiles.

Strong ROC discrimination across most tauopathy and inclusion classes

Receiver operating characteristic (ROC) curves for each category are shown in Fig. 2c, with corresponding area under the curve (AUC) values reported in Table 2. In the low-magnification model, AUC values ranged from 0.872 to 0.989. AD showed the highest discrimination, with an AUC of 0.989, while GGT showed the lowest at 0.872 and CTE was also comparatively lower, although still favourable, at 0.894. The remaining groups fell between 0.903 and 0.972. In the high-magnification model, there was a wider range of AUC values, from 0.601 to 0.989. The highest AUC was observed for GFA at 0.989, with all other inclusion classes at or above 0.925 with the exception of PiD-associated RA, which showed moderate discrimination at 0.770, and GGT-associated GAI, which showed the weakest at 0.601. Several groups showed favourable AUC values despite lower recall at the final multigroup prediction threshold; PiD and CTE, for example, remained distinguishable from other disease groups across thresholds, although many images were assigned to another group when the highest-probability prediction was selected.

Discussion

This study supports the feasibility of using transfer learning through a general-purpose computer vision platform to classify neurodegenerative disease-associated tau pathology at both regional (low magnification) and cellular (high magnification) levels. The low-magnification classifier was more accurate on individual images, whereas the high-magnification classifier was more accurate when each case was assigned to the category predicted for most of its images. Performance varied substantially, with lower classification accuracy for rare, underrepresented disease groups, or inclusions with morphologically overlapping features. Conversely, the models showed high recall for several diagnostically important categories, including AD at low and high magnification, NFTs associated with

AD, GFAs associated with ARTAG, and TAs and APs characteristic of PSP and CBD, respectively. These findings indicate that openly accessible and computationally lightweight approaches can extract meaningful information from pTau-immunostained tissue sections without requiring custom model development or specialised computational infrastructure. They also show that classification performance depends on both image magnification and the amount of surrounding histological context available to the model.

The greater image-level prediction accuracy of the low-magnification model may reflect the additional spatial and tissue-level context available within each image. At 5× magnification, the model captured broader histological patterns, including anatomical distribution, overall pTau pathology burden, cortical architecture, and the coexistence of different inclusion types. These broader histological features form part of the integrated assessment of tau pathology used in routine neuropathological assessment[1,17]. By contrast, the high-magnification model classified individual inclusions in isolation and therefore relied on subtler differences in cellular shape and staining distribution, features that may overlap across tauopathies or exist along a morphological continuum. Images were also resized by the platform to 224 × 224 pixels, which may have reduced visibility of fine cellular processes and other subtle morphological features relevant to distinguishing certain inclusion types. At case level, the ranking of the two models was reversed: the high-magnification model classified cases more accurately than the low-magnification model, despite its lower accuracy on individual images. However, the two models differed not only in magnification but also in classification task, class composition, and the number of images available per case, both for training and for case-level aggregation. Their difference in performance should therefore not be attributed to magnification alone.

The lower-performing categories illustrate the combined effects of biological overlap and limited dataset representation. At low magnification, GGT was frequently confused with other 4R tauopathies, particularly PSP and CBD, likely reflecting similarities in astrocytic inclusion morphologies[3,4,18]. At low magnification, PiD was most frequently misclassified as AD, likely reflecting the similarly dense pancortical tau pathology seen in both disorders at this image scale, with concurrent AD pathology in some PiD cases potentially contributing further[19]. An additional confounding factor for CTE classification at low magnification is regional sampling. A diagnostic feature of CTE is perivascular astrocytic pTau pathology in the form of TSAs at the depths of cortical sulci, and images that did not consistently capture this area would lack a pattern on which the pathological diagnosis depends[20]. Similar challenges were evident at high magnification, where RAs associated with PiD and GAIs characteristic of GGT were often assigned to morphologically similar astrocytic classes. The TSA class frequently captured misclassified GAI and RA images, resulting in high recall but lower precision, consistent with the morphological overlap between CTE and ARTAG-associated TSAs and other astrocytic pTau inclusions[21]. Classification of isolated inclusions removes much of the spatial context used in traditional specialist neuropathological assessment, making subtle or continuous morphological differences more difficult to resolve for computational models. These difficulties were compounded by the rarity of CTE and GGT cases, which limited the number of independent cases available for training

and consequently, the range of biological variation represented in the dataset. Although augmentation increased the number and visual diversity of images, it could not substitute for broader case-level representation, particularly for rare diseases. Notably, several groups with low recall such as PiD and CTE retained favourable ROC discrimination. This indicates that discriminative features were learned but were outweighed at the point of multigroup assignment. A further constraint arises where group boundaries are not defined by pTau pathology alone. The distinction between PART and low ADNC, and between low ADNC and high ADNC-associated NPs, depends on the presence and regional extent of concurrent amyloid-β pathology, which cannot be assessed on pTau immunohistochemistry alone[5,22]. The improved classification when PART was grouped with Low AD at low magnification, and ageing NP from cases with low level ADNC with AD NP at high magnification, reflects an inherent limitation of pTau-immunostained images, which do not provide the additional amyloid-β information required to distinguish these categories, rather than a failure of the trained classifier.

Our findings should be considered in the context of previous computational approaches to tau pathology assessment, which have ranged from narrowly defined classification of selected inclusion types to broader assessment of overall pathological burden. Google AutoML, a proprietary code-free image classification platform, distinguished PSP-associated TAs, CBD-associated APs and AD pathology associated NPs on pTau-immunostained sections with 100% recall for TAs and NPs and 89% for APs in a 300-image test set from 60 patients[23]. A general-purpose vision–language model, GPT-4Vision, classified the same three lesion types with 90% accuracy from 20 examples supplied in the prompt, matching a CNN trained on 100 labelled images[24]. Other studies have used custom-built models for quantification of overall tau pathology burden or stage, an approach requiring programming expertise and dedicated computational resources. CNNs have been developed for detection and staging of AD-associated NFTs across whole-slide images[25], and for quantification of NFT burden, outperforming Braak NFT staging in predicting antemortem cognitive impairment[26]. Weakly supervised and foundation-model approaches can capture broader spatial patterns without inclusion-level annotation, but generally require much greater computational resources and may offer limited interpretability[27].

Strengths of this study include the breath of tau neuropathology represented, spanning multiple disease groups and a range of characteristic neuronal and astrocytic inclusion morphologies, together with assessment of two complementary spatial scales. Standardised image acquisition and expert-reviewed annotations provided consistent neuropathological ground truth, while evaluation at both image and case level enabled assessment of how classification performance changed when information from images was considered together. The use of an accessible, code-free workflow, and development of an open access web-based application further support reproducibility and potential wider use of the trained classifiers.

Despite these strengths, several limitations should be considered. The dataset remained relatively small at the case level, particularly for rare groups such as CTE and GGT, limiting representation of the full spectrum of within-disease variation. Although image augmentation increased image-level diversity, it

could not substitute for additional independent cases and, where applied to test images, did not increase biological independence. The study also focused predominantly on cortical regions, and the extent to which the models generalise to tau pathology in subcortical regions remains uncertain. In addition, all tissue was derived from a single centre and processed using a single immunostaining protocol and digital imaging platform, which may also limit generalisability to material prepared under different technical conditions. External validation across independent cohorts and platforms will therefore be important before implementation in diagnostic practice. A further practical limitation was the time required for manual selection of regions of interest. These limitations define clear priorities for future work. Multicentre validation using larger and more diverse case cohorts, particularly for rare tauopathies, should assess robustness across different staining protocols, digital slide scanners and anatomical regions. Combining region-level contextual information with inclusion-level morphology may further improve classification, while the addition of localisation, segmentation and explainability methods could support more interpretable analysis of whole-slide digital images. Direct comparison with task-specific networks and emerging pathology foundation models will also be important in defining the advantages and limitations of this lightweight approach. Overall, this study demonstrates that a general-purpose vision model can capture meaningful disease-associated tau patterns and provides an accessible benchmark for the development of more advanced morphology-based tools in neuropathology.

Methods

Study cohort and neuropathological examination

All cases were obtained from the UCL Queen Square Brain Bank for Neurological Disorders (QSBB). Pathological diagnoses for all cases included in the study were confirmed by neuropathologist (Z.J.) with expertise in neurodegenerative diseases, using applicable consensus diagnostic and staging criteria[18,22,28–30]. Tissue collection, processing, and diagnostic assessment were performed as part of routine brain bank protocols. The QSBB operates under ethical approval from the NHS Health Research Authority London–Central Research Ethics Committee (REC 23/LO/0044), and the tissue used in this study was provided under this approval. Written informed consent for the use of donated tissue in research was obtained from all donors or their next of kin. All methods were performed in accordance with the relevant guidelines and regulations. Archival 7μm formalin-fixed paraffin-embedded tissue sections immunostained for phosphorylated tau (AT8; MN1020; Thermo Scientific; 1:1,200) using automated platforms (Menarini Diagnostics or Ventana, Roche) were used for this study. Detection employed horseradish peroxidase with diaminobenzidine chromogen and haematoxylin counterstain, with appropriate controls included in each run. The anatomical regions used for each disease group are outlined in Table 1. pTau immunostained glass slides were digitised using a Hamamatsu NanoZoomer S360 scanner at 40× (0.23 μm/pixel; JPEG compression 70) and whole-slide images were accessed via NZConnect.

Image sampling and dataset construction

To standardise image sampling, non-overlapping annotation guide boxes were manually placed across the grey matter, as outlined in Table 1, at two different magnifications. The low magnification set comprised 4 mm² guide boxes sampled at 5× magnification and downsampled to fixed dimensions of 900 × 900 pixels. These images were designed to capture low-resolution, pancortical patterns of pTau pathology across the eight disease groups. The high-magnification set comprised variable-sized guide boxes placed at 40× to isolate individual cells with inclusion pathology. Nine morphological classes were represented: RA, TA, AP, GAI, TSA, GFA, ageing NP, AD NP and NFT. All annotations were reviewed by a neuropathologist (Z.J) to ensure consistency of image selection and disease group or morphological class assignment. Images were exported in PNG format using standardised dimensions and a neutral gamma value of 1.8. The number of images contributing to each category in the training and testing datasets are summarised in Fig. 1. Training and testing images were separated at the case-level, with no cases being shared across datasets to ensure true generalisation during evaluation.

Image augmentation

Data augmentation was applied to the training images for the high-magnification model to increase image diversity and reduce overfitting. Each source image was augmented using four transformations, including rotations of 90° and 270° and increases in the intensity of the red and blue colour channels. These transformations were selected to simulate variation in tissue orientation, staining and image acquisition across laboratory settings. Augmentation increased the number of training images in each pTau inclusion class from 1,800 to 9,000. For GFA and ageing NP classes, where fewer test images were available because of limited case numbers containing relevant morphological features, rotational transformations of 90°, 180° or 270° were performed to increase sample size. This process was repeated until both of these inclusion classes achieved 400 images.

Model training and deployment

GTM, a browser-based image-classification platform, was used to develop CNN models with minimal computational requirements and without custom model architecture design. GTM applies transfer learning using a pretrained CNN to support classification of domain-specific image datasets. Two models were trained using digitised pTau-immunostained post-mortem brain tissue sections prepared on glass slides: a low-magnification model for classifying tauopathies from 5× images measuring 900 × 900 pixels, and a high-magnification model for classifying specific tau-related inclusion morphologies from 40× images. Models were trained through the GTM browser interface; the low-magnification model used the default hyperparameters of 50 epochs, a batch size of 16 and a learning rate of 0.001, while the high-magnification model used slightly modified hyperparameters (25 epochs; batch size of 16; and learning rate of 0.0005) to prevent overfitting. Uploaded images were automatically partitioned by the platform into training (85%) and internal validation (15%) subsets, with no manual control over this split available within the platform. Training and validation accuracy and loss were monitored via the platform's built-in "Under the hood" interface. The number of training images in the low-magnification model varied between disease groups according to case and image availability, as shown in Fig. 1. The high-magnification model included 1,800 original images per morphological class, increasing to 9,000

images per class following the augmentation procedure described above. Trained models were exported in TensorFlow.js format as a model.json file, a metadata.json file and binary weight files to enable batch inference without modifying the trained model weights, the exported models were integrated into a custom HTML/JavaScript web application hosted on Netlify. The application reproduced the prediction outputs of GTM while enabling folder-level image uploads. The developed classifiers through web-based application are publicly available at https://tau-net.org/.

Statistical analysis

Statistical analyses were performed in R version 4.5.0. Exported predicted probabilities for disease and individual inclusion classes were analysed using the caret, pROC, Metrics, ggplot2, and patchwork packages. Each image was assigned to the disease group or inclusion class with the highest predicted probability. Case-level accuracy was calculated by aggregating image-level predictions within each case, with each case assigned to the category receiving the majority of predictions across its constituent images. Class-specific performance was evaluated using precision, recall, F1 score, and one-versus-rest area under the receiver operating characteristic curve (AUC). Precision represents the proportion of predictions assigned to a class that were correct, whereas recall represents the proportion of images belonging to that class that were correctly identified. Confusion matrices and ROC curves were generated to assess classification performance and patterns of misclassification between distinct tauopathy and individual inclusion classes. Large language models (ChatGPT, OpenAI and Claude, Anthropic) were used to assist in generating code for statistical analyses, data visualisation and creating the HTML/JavaScript web application. All code, statistical tests, and outputs were manually reviewed and validated by the authors to ensure methodological appropriateness and accuracy of results.

Declarations

Competing interests

The authors report no competing interests.

Funding

The authors received no specific funding for this work.

Author Contribution

L.S., A.C., Z.J. and P.W.C contributed equally to the study. L.S., A.C., Z.J., and P.W.C., conceived and designed the study. L.S. and A.C. developed the image analysis workflow, trained and evaluated the models and performed the statistical analyses. L.S., A.C, Z.J and P.W.C. prepared the figures, and wrote the manuscript. S.S., C.V., G.C., T.C., Z.F., M.B., N.C., H.R.N and Y.M.L. contributed to image acquisition,

annotation, and data curation. Z.J. and P.W.C. provided supervision and project oversight. All authors critically reviewed the manuscript and approved the final version.

Acknowledgement

We thank the patients and their families, without whose generous donations this study would not have been possible. We also thank the QSBB administrative and technical staff for assistance in tissue preparation. Large language models (ChatGPT, OpenAI and Claude, Anthropic) were used to improve the clarity and conciseness of specific phrases and sentences in the manuscript. All suggestions were critically reviewed and revised by the authors to ensure the accuracy and integrity of the scientific content.

Data Availability

Labelled histopathology images used for model training and testing may be shared upon reasonable request by qualified investigators for purposes of replicating procedures and results, if in compliance with ethical approval. Trained model weights and custom JavaScript code used for the web application are openly available in a public GitHub repository (https://github.com/JaunmuktaneLab).

References

1. Kovacs, G. G., Ghetti, B. & Goedert, M. Classification of diseases with accumulation of Tau protein. *Neuropathol. Appl. Neurobiol.* **48**, e12792. https://doi.org/10.1111/nan.12792 (2022).
2. Kovacs, G. G. Invited review: Neuropathology of tauopathies: principles and practice. *Neuropathol. Appl. Neurobiol.* **41**, 3–23. https://doi.org/10.1111/nan.12208 (2015).
3. Shi, Y. et al. Structure-based classification of tauopathies. *Nature* **598**, 359–363. https://doi.org/10.1038/s41586-021-03911-7 (2021).
4. Ahmed, Z. et al. Globular glial tauopathies (GGT): consensus recommendations. *Acta Neuropathol.* **126**, 537–544. https://doi.org/10.1007/s00401-013-1171-0 (2013).
5. Crary, J. F. et al. Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol.* **128**, 755–766. https://doi.org/10.1007/s00401-014-1349-0 (2014).
6. Bieniek, K. F. et al. The Second NINDS/NIBIB Consensus Meeting to Define Neuropathological Criteria for the Diagnosis of Chronic Traumatic Encephalopathy. *J. Neuropathol. Exp. Neurol.* **80**, 210–219. https://doi.org/10.1093/jnen/nlab001 (2021).
7. Julian, D. R., Bahramy, A., Neal, M., Pearce, T. M. & Kofler, J. Current Advancements in Digital Neuropathology and Machine Learning for the Study of Neurodegenerative Diseases. *Am. J. Pathol.* **195**, 2102–2117. https://doi.org/10.1016/j.ajpath.2024.12.018 (2025).

8. Nelvagal, H. R. et al. Quantitative pathology and APOE genotype reveal dementia risk and progression in Lewy body disease. *Brain* https://doi.org/10.1093/brain/awag114 (2026).
9. Cullinane, P. W. et al. Data-driven modelling of tau pathology reveals distinct progressive supranuclear palsy subtypes. *Brain* https://doi.org/10.1093/brain/awag131 (2026).
10. Robinson, J. L. et al. Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain* **141**, 2181–2193. https://doi.org/10.1093/brain/awy146 (2018).
11. Carney, M. et al. Teachable Machine: approachable web-based tool for exploring machine learning classification in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* 1–8 (Association for Computing Machinery, 2020). 1–8 (Association for Computing Machinery, 2020). (2020). https://doi.org/10.1145/3334480.338283
12. Deng, J. et al. ImageNet: a large-scale hierarchical image database in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009). https://doi.org/10.1109/CVPR.2009.5206848
13. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. C. MobileNetV2: inverted residuals and linear bottlenecks in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 4510–4520 (IEEE, 2018). 4510–4520 (IEEE, 2018). (2018). https://doi.org/10.1109/CVPR.2018.00474
14. Forchhammer, S., Abu-Ghazaleh, A., Metzler, G., Garbe, C. & Eigentler, T. Development of an Image Analysis-Based Prognosis Score Using Google's Teachable Machine in Melanoma. *Cancers (Basel)*. 14. https://doi.org/10.3390/cancers14092243 (2022).
15. Singla, N. et al. A Pilot Study of Breast Cancer Histopathological Image Classification Using Google Teachable Machine: A No-Code Artificial Intelligence Approach. *Cureus* **17**, e87301. https://doi.org/10.7759/cureus.87301 (2025).
16. Jungo, P. & Hewer, E. Code-free machine learning for classification of central nervous system histopathology images. *J. Neuropathol. Exp. Neurol.* **82**, 221–230. https://doi.org/10.1093/jnen/nlac131 (2023).
17. Cullinane, P. W. et al. Pathology of neurodegenerative disease for the general neurologist. *Pract. Neurol.* **24**, 188–199. https://doi.org/10.1136/pn-2023-003988 (2024).
18. Kovacs, G. G. et al. Distribution patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathol.* **140**, 99–119. https://doi.org/10.1007/s00401-020-02158-2 (2020).
19. Hof, P. R., Bouras, C., Perl, D. P. & Morrison, J. H. Quantitative neuropathologic analysis of Pick's disease cases: cortical distribution of Pick bodies and coexistence with Alzheimer's disease. *Acta Neuropathol.* **87**, 115–124. https://doi.org/10.1007/bf00296179 (1994).
20. McKee, A. C. et al. The first NINDS/NIBIB consensus meeting to define neuropathological criteria for the diagnosis of chronic traumatic encephalopathy. *Acta Neuropathol.* **131**, 75–86. https://doi.org/10.1007/s00401-015-1515-z (2016).
21. Kovacs, G. G. et al. Aging-related tau astrogliopathy (ARTAG): harmonized evaluation strategy. *Acta Neuropathol.* **131**, 87–102. https://doi.org/10.1007/s00401-015-1509-x (2016).

22. Hyman, B. T. et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* **8**, 1–13. https://doi.org/10.1016/j.jalz.2011.10.007 (2012).
23. Koga, S., Ghayal, N. B. & Dickson, D. W. Deep Learning-Based Image Classification in Differentiating Tufted Astrocytes, Astrocytic Plaques, and Neuritic Plaques. *J. Neuropathol. Exp. Neurol.* **80**, 306–312. https://doi.org/10.1093/jnen/nlab005 (2021).
24. Ono, D., Dickson, D. W. & Koga, S. Evaluating the efficacy of few-shot learning for GPT-4Vision in neurodegenerative disease histopathology: A comparative analysis with convolutional neural network model. *Neuropathol. Appl. Neurobiol.* **50**, e12997. https://doi.org/10.1111/nan.12997 (2024).
25. Vizcarra, J. C. et al. Toward a generalizable machine learning workflow for neurodegenerative disease staging with focus on neurofibrillary tangles. *Acta Neuropathol. Commun.* **11**, 202. https://doi.org/10.1186/s40478-023-01691-x (2023).
26. Marx, G. A. et al. Artificial intelligence-derived neurofibrillary tangle burden is associated with antemortem cognitive impairment. *Acta Neuropathol. Commun.* **10**, 157. https://doi.org/10.1186/s40478-022-01457-x (2022).
27. Verma, R. et al. Domain-specific foundation model improves AI-based analysis of neuropathology. Preprint at arXiv (2025). https://doi.org/10.48550/arXiv.2512.05993
28. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259. https://doi.org/10.1007/bf00308809 (1991).
29. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791–1800. https://doi.org/10.1212/wnl.58.12.1791 (2002).
30. Braak, H. et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging*. **24**, 197–211. https://doi.org/10.1016/s0197-4580(02)00065-9 (2003).

Figures



Figure 1

Overview of training and testing datasets of the low- and high-magnification tauopathy classification models. For the low-magnification model, each disease group label is accompanied by the number of cases (top row). The number beneath each representative image represents the total number of images extracted for that disease group. Images were derived from AT8-immunostained digital whole slide images at 5x magnification. For the high-magnification model, inclusion class labels are similarly

annotated with the number of cases. Representative images illustrate the morphological features of each class. Image augmentation (including rotations and colour variations) are schematically represented. The numbers beneath each image indicate the total number of images used for testing and training following augmentation. PiD, Pick's disease; PSP, progressive supranuclear palsy; CBD, corticobasal degeneration; GGT, globular glial tauopathy; CTE, chronic traumatic encephalopathy; PART, primary age-related tauopathy; AD, Alzheimer's disease; NP, neuritic plaque; RA, ramified astrocyte; TA, tufted astrocyte; AP, astrocytic plaque; GAI, globular astrocytic inclusion; TSA, thorn-shaped astrocyte; GFA, granular/fuzzy astrocyte; NFT, neurofibrillary tangle.

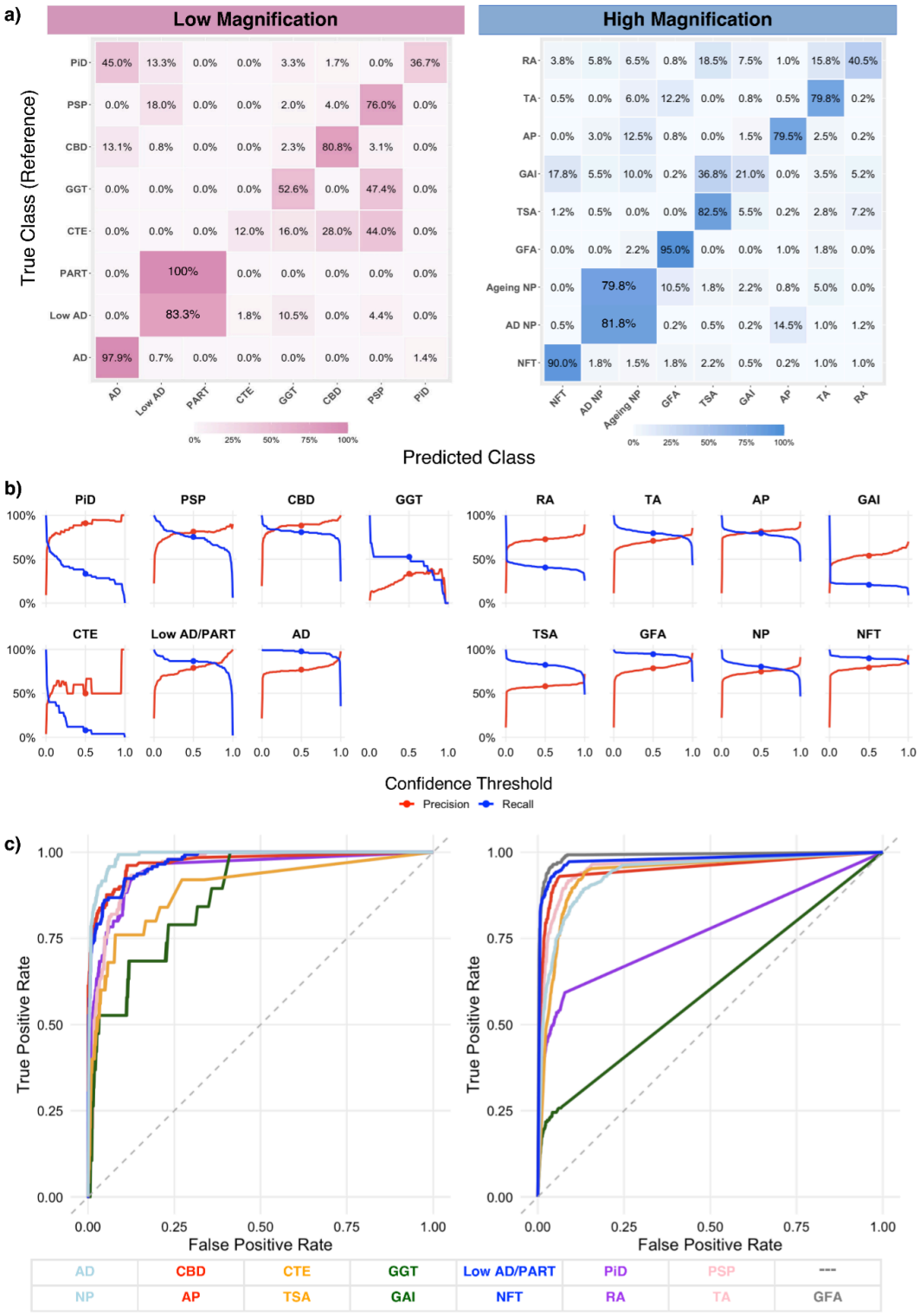


Figure 2

Performance of low- and high-magnification tauopathy classification models. (a) Confusion matrices for low-magnification (5×; left) and high-magnification individual inclusion (40×; right) models. Rows represent the true class (reference) and columns represent the predicted class. Values indicate the percentage of cases assigned to each group. (b) Precision (red) and recall (blue) plotted against classification confidence threshold for each category. Panels are shown separately for low-magnification

disease-level predictions (left) and high-magnification inclusion-level predictions (right). Dots on each curve indicate performance at the default threshold of 0.5. (c) One-vs-rest receiver operating characteristic (ROC) curves for each class for low-magnification (left) and high-magnification (right) models. The dashed diagonal line indicates chance performance. Colours correspond to the class labels shown below the graph area.

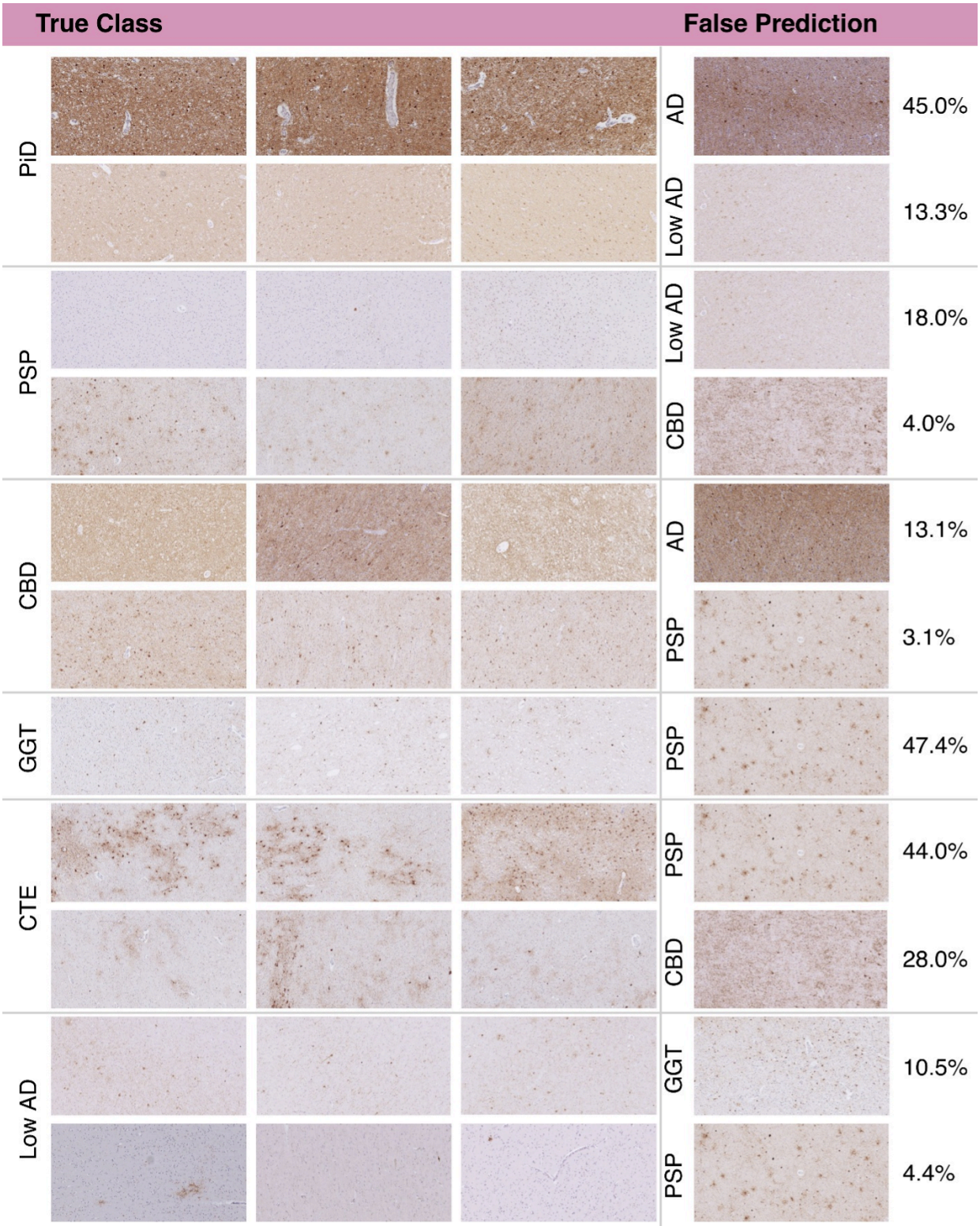


Figure 3

Representative examples of misclassified low-magnification images. For each true diagnostic class (left), three example images are shown that were incorrectly classified by the model. The false predicted classes shown correspond to the two most frequent incorrect predictions for each diagnosis, where the proportion exceeded 2.5% of predictions. The percentage displayed alongside each incorrect prediction indicates the proportion of images from that true class assigned to predicted class. The single image displayed in the right column represents a correctly classified reference example from the predicted class. Images are derived from low-magnification (5×) AT8-immunostained digital whole slide images of frontal cortex from distinct tauopathies.

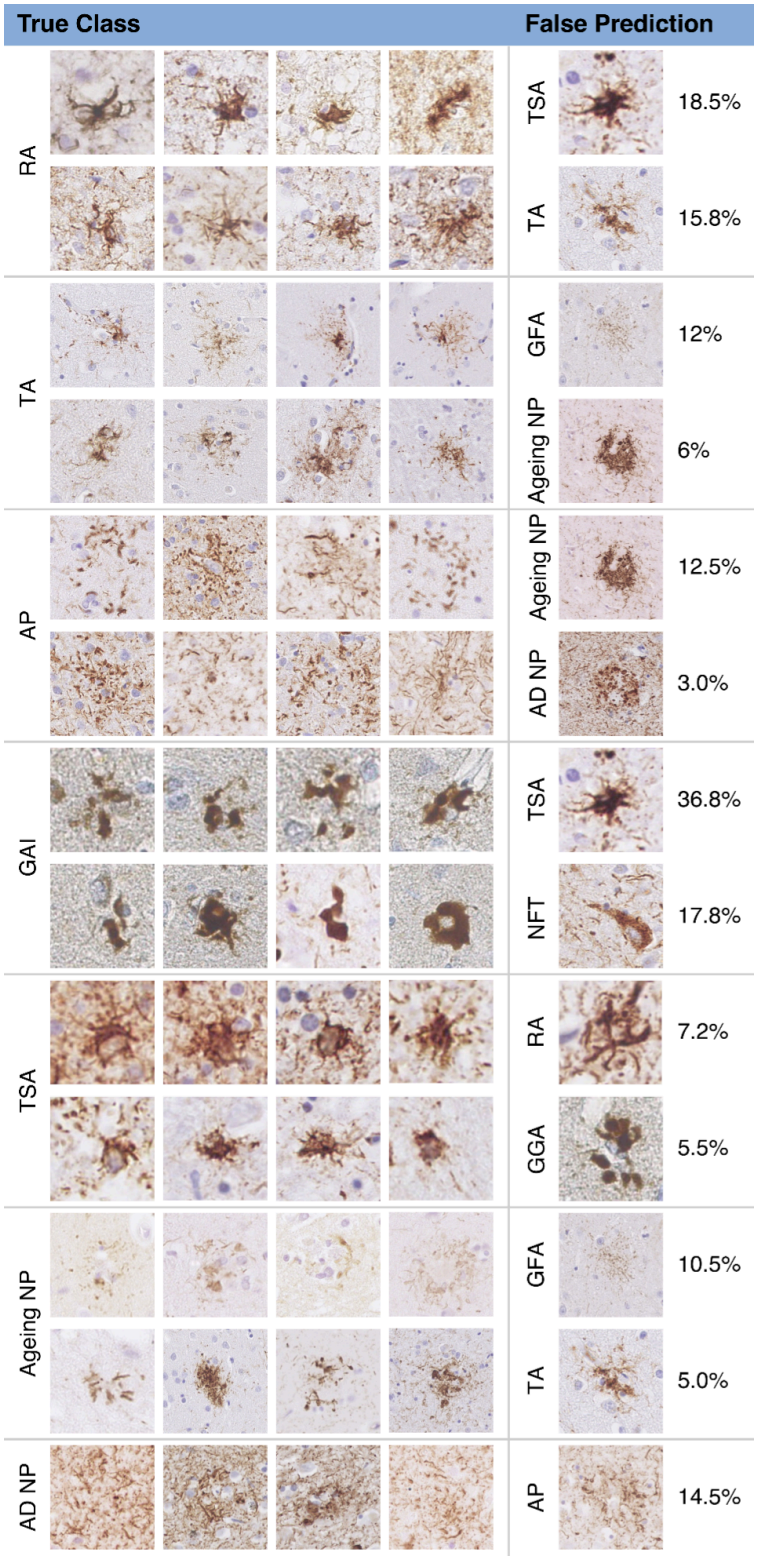


Figure 4

Representative examples of misclassified high-magnification images. For each true inclusion class (left), four example images are shown that were incorrectly classified by the model. The false predicted classes shown correspond to the two most frequent incorrect predictions for each class, where the proportion exceeded 2.5% of predictions. The percentage displayed alongside each incorrect prediction indicates the proportion of images from the true class assigned to the predicted class. The single image

displayed in the right column represents a correctly classified reference example from the predicted class. Images are derived from high-magnification (40×) AT8-immunostained digital whole slide images showing individual cellular pTau inclusions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.pdf