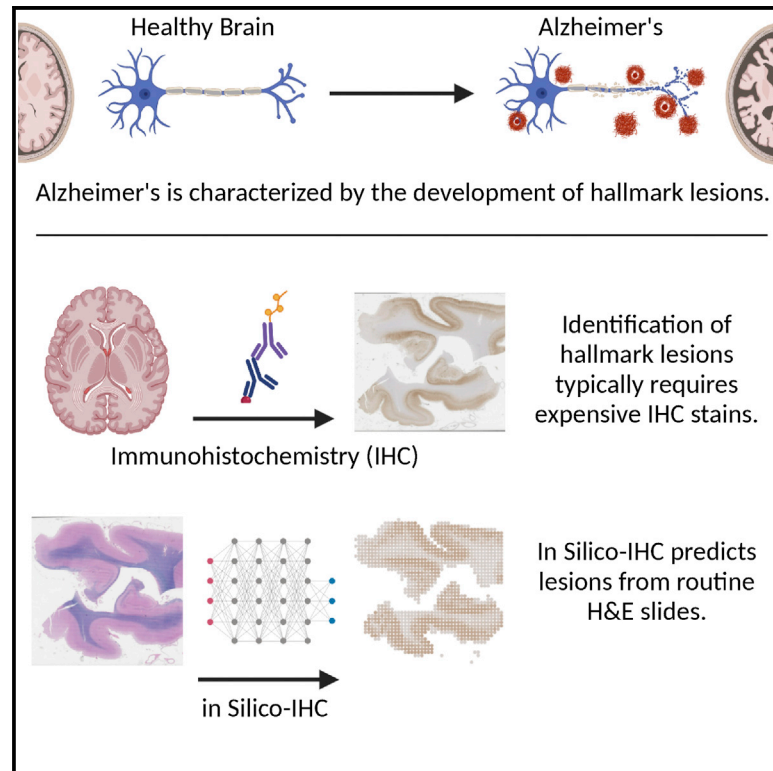


AI-enabled *in silico* immunohistochemical characterization for Alzheimer's disease

Graphical abstract



Authors

Bryan He, Syed Bukhari, Edward Fox, ..., Maria Corrada, Thomas Montine, James Zou

Correspondence

tmontine@stanford.edu (T.M.), jamesz@stanford.edu (J.Z.)

In brief

He et al. develop *in silico* IHC, a deep learning-based method that uses routine histochemical stains to identify features that conventionally require more expensive immunohistochemical stains. The method is applied to Alzheimer's disease, where it identifies several hallmark changes that characterize progression of the disease.

Highlights

- Deep learning-based computational processing of routine histochemical samples
- Identification of features conventionally requiring immunohistochemical stains
- Apply *in silico* IHC to identify hallmark changes of Alzheimer's disease



Article

AI-enabled *in silico* immunohistochemical characterization for Alzheimer's disease

Bryan He,¹ Syed Bukhari,² Edward Fox,² Abubakar Abid,³ Jeanne Shen,² Claudia Kawas,⁴ Maria Corrada,⁴ Thomas Montine,^{2,*} and James Zou^{1,2,5,6,7,*}

¹Department of Computer Science, Stanford University, Stanford, CA 94305, USA

²Department of Pathology, Stanford University, Stanford, CA 94305, USA

³Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

⁴Department of Neurology, University of California, Irvine, Irvine, CA 92697, USA

⁵Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA

⁶Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA

⁷Lead contact

*Correspondence: tmontine@stanford.edu (T.M.), jamesz@stanford.edu (J.Z.)

<https://doi.org/10.1016/j.crmeth.2022.100191>

MOTIVATION Immunohistochemistry (IHC) is used widely across diseases to identify proteins of interest and aid disease diagnosis. However, IHC is time consuming and expensive to prepare at a large scale. For evaluating the progression of Alzheimer's disease (AD), several hallmark changes must be identified via IHC staining across many regions of the brain. The cost of preparing many samples with IHC staining for AD has been a major barrier to large genomic-endophenotype studies. As a result, research and diagnosis of AD and other diseases would benefit from the ability to identify these proteins without the need for IHC.

SUMMARY

We develop a deep learning approach, *in silico* immunohistochemistry (IHC), which takes routinely collected histochemical-stained samples as input and computationally generates virtual IHC slide images. We apply *in silico* IHC to Alzheimer's disease samples, where several hallmark changes are conventionally identified using IHC staining across many regions of the brain. *In silico* IHC computationally identifies neurofibrillary tangles, β -amyloid plaques, and neuritic plaques at a high spatial resolution directly from the histochemical images, with areas under the receiver operating characteristic curve of between 0.88 and 0.92. *In silico* IHC learns to identify subtle cellular morphologies associated with these lesions and can generate *in silico* IHC slides that capture key features of the actual IHC.

INTRODUCTION

Cellular morphology is closely linked to tissue function and disease diagnosis. A common tool in pathology for assisting with disease diagnosis is immunohistochemical (IHC) staining, which is used to identify specific proteins of interest in a tissue. In this work, we propose to use deep learning to computationally generate *in silico* IHC staining. We demonstrate that deep learning algorithms can identify subtle features in cellular morphology, which are associated with diseases and previously required IHC to visualize. These results also open the door for computational approaches to potentially reduce the need of performing time-consuming or expensive experimental IHC staining.

In the standard workflow, tissue samples collected for pathological diagnosis have a section prepared with a hematoxylin and

eosin (H&E) stain for general histologic assessment. Specialized IHC stains are additionally applied to other sections of the same tissue to identify structures or specific molecules that are difficult to directly observe in the H&E-stained sample. These IHC-stained slides are commonly studied by eye under a microscope. Digitization of these slides into whole-slide images (WSIs) now allows for computational assistance with evaluating slides (Fu et al., 2020; He et al., 2020; Kather et al., 2020).

Performing all necessary IHC stains on a sample can cost hundreds of dollars and requires several days to process, which can be avoided by using the *in silico* IHC. Additionally, the computationally generated stain would allow the IHC to be run on the same section of tissue as the original H&E-stained slide, rather than a different section, and removes artifacts that appear on the real IHC stains. In addition to these advantages in the diagnostic setting, *in silico* IHC has the potential to make major



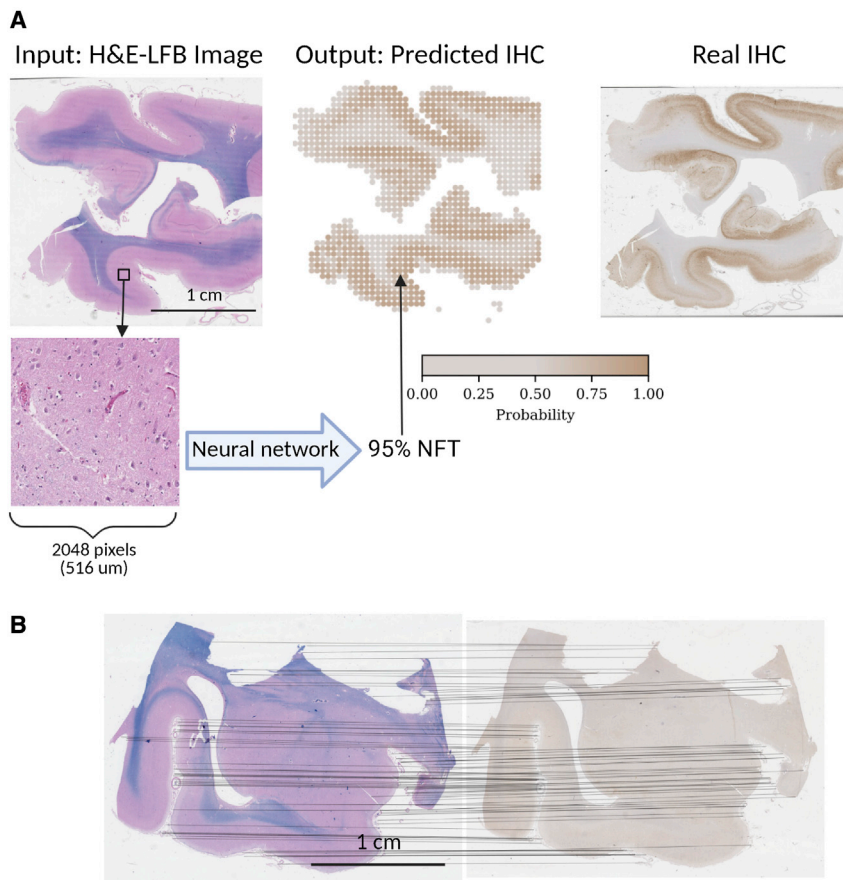


Figure 1. *In silico* IHC

(A) Procedure for *in silico* staining an H&E-LFB WSI. The H&E-LFB image is divided into patches, which are given to the trained neural network to separately predict the presence of each hallmark change, which are extracted from the corresponding IHC patch. The predictions are then combined into a synthetic IHC image. The corresponding real IHC-stained slide is shown for comparison.

(B) Example showing the registration process, which is used to provide supervision to train *in silico* IHC. Key points from each of the sections are detected, and the matching key points are indicated by lines.

focused on IHC targeting specific cell types (Xu et al., 2016; Sharma et al., 2017; Jackson et al., 2020; Liu et al., 2021), such as neoplastic or necrotic cells, which are more visually distinct on H&E slides than the AD lesions we studied.

We present *in silico* IHC, a system for our *in silico* IHC staining process (Figure 1). As a proof of concept, we apply *in silico* IHC to AD. The brains of patients with AD have several hallmark neuropathologic lesions: β -amyloid ($A\beta$) plaques, neurofibrillary tangles (NFT), and neuritic plaques (Hyman et al., 2012; Montine et al., 2012). These hallmark changes typically occur in specific regions of the brain before the onset of cognitive impairment, and then increase in density and distribution as the disease ad-

contributions to genomic research that relies on IHC-generated phenotypes. For example, large genetic association studies of Alzheimer's disease (AD) neuropathologic endophenotypes have been severely limited by the lack of IHC data on research autopsy brains (Beecham et al., 2014).

Recent advances in deep learning have produced increasingly accurate image recognition models (Deng et al., 2009; Krizhevsky et al., 2012; Huang et al., 2017). These advances have resulted in deep learning being applied across medicine fields (Esteva et al., 2017; Ouyang et al., 2020). Within pathology, deep learning has been used to classify disease subtypes and predict mutations (Campanella et al., 2019; Fu et al., 2020; Kather et al., 2020; Lu et al., 2021) and to interpret IHC stains (Signaevsky et al., 2019; Koga et al., 2021). Combined with spatial transcriptomics, deep learning has also been used to link cell morphologic features with localized gene expressions (He et al., 2020; Levy-Jurgenson et al., 2020). Finally, deep learning has been used to transform unstained samples into virtual H&E stains (Rivenson et al., 2019) and to label cellular constituents, such as the nuclei and membrane, from microscopy images (Christiansen et al., 2018). H&E and IHC are not commonly prepared on the same tissue section, making supervised learning more difficult. As a result, computationally generating IHC stains directly from H&E images has been less explored. Studies that have used IHC slides have typically

vances. IHC staining is used to highlight instances of each of these hallmark changes. IHC for $A\beta$ is used to highlight instances of $A\beta$ plaques, and IHC for pathologic forms of tau (often collectively called phospho-tau) are used to highlight instances of NFTs and NPs, which can be differentiated via visual inspection. The IHC assessments are used in consensus pathologic evaluation to determine the regional distribution of $A\beta$ plaques, regional distribution of NFTs, and regional density of NPs. Together, these form the basis of the current National Institutes on Aging–Alzheimer's Association consensus guidelines for the neuropathologic assessment of AD (Hyman et al., 2012; Montine et al., 2012).

To train and evaluate our system, we collected a dataset consisting of brain autopsies from a total of 160 patients, consisting of 704 samples from different regions of the brain. Within a single patient, the presence of hallmark changes may vary from region to region, resulting in the need for multiple samples for each patient. Each sample is divided into sections for staining. One section from all samples is stained with H&E combined with Luxol fast blue (LFB), which is commonly used to highlight the white matter in the brain. This combined stain is referred to as H&E-LFB. Additionally, separate sections from the sample are prepared with $A\beta$ and pathologic tau IHC stains. The regions and stains used in our study follow the recommendations of the National Institute on Aging–Alzheimer's Association guidelines for assessing neuropathologic change in AD (Hyman et al., 2012; Montine et al., 2012).

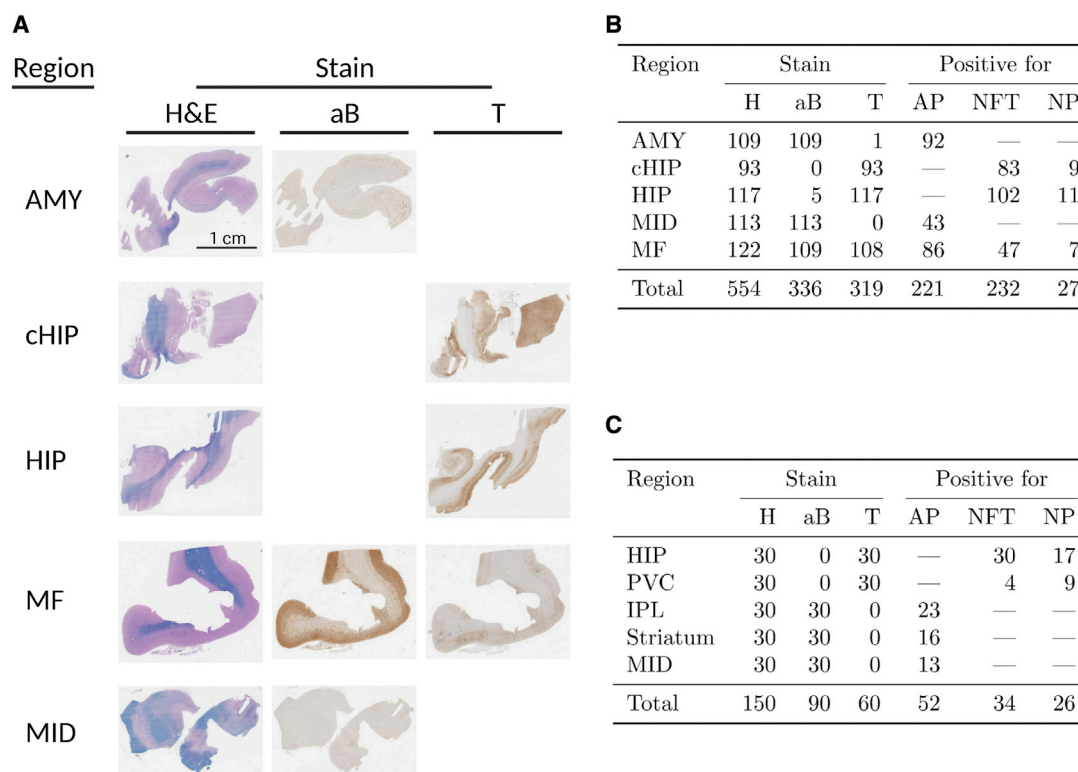


Figure 2. Overview of the dataset

(A) WSIs for an example patient. The samples are collected from five regions of the brain (amygdala [AMY], contralateral hippocampus [cHIP], hippocampus [HIP], midbrain [MID], and middle frontal gyrus [MF]). All samples have an H&E-LFB (H) slide available, along with at least one slide of amyloid-Beta (aB) or phosphorylated Tau (T).

(B and C) The number of slides available for each region of the brain and number of slides with positivity for amyloid plaques (AP) NFTs and neuritic plaques for (B) training and testing the model and (C) additional evaluation (additionally includes samples from the primary visual cortex [PVC], inferior parietal lobule [IPL], and striatum). The samples were prepared for determining the level of AD neuropathologic change, and only the recommended brain regions to be evaluated were stained for most patients (Hyman et al., 2012; Montine et al., 2012).

We then divide the dataset into separate training, validation, and test sets by patient. Training a deep learning model requires a dataset consisting of pairs of input and expected output (e.g., H&E-LFB images and presence of each hallmark change). Typically, the expected output for the dataset is generated by manual annotation (Litjens et al., 2018) or by combining slide-level annotations and multiple-instance learning (Campanella et al., 2019; Lu et al., 2021). However, we can computationally align the serial slides prepared with different stains to provide annotation for the H&E-LFB images. This approach reduces the need for manual annotation. After training, *in silico* IHC achieved areas under the receiver operating characteristic curve (AUROCs) of 0.91 (95% CI, 0.88–0.95) for classifying the presence of NFTs, 0.92 (0.87–0.94) for neuritic plaques, and 0.88 (95% CI, 0.82–0.93) for A β plaques on the held-out test set.

RESULTS

Data curation

Our dataset consists of autopsied brains from 160 patients. Each brain is divided into several regions of interest, resulting in 704 samples collected from multiple regions of brain (Figure 2A).

The tissue sample from each region is prepared into a formalin-fixed paraffin-embedded block. These blocks are cut into serial sections of five μ m thickness, with one slide prepared with H&E-LFB staining, along with at least one of pathologic tau and A β staining. We split the dataset into a training set consisting of 91 patients, a validation of 20 patients, and a testing set of 19 patients (Figure 2B). As an additional test, data for 30 consecutive patients were collected at a different time with all sections necessary for a full analysis of the level of neuropathologic change in each brain (Figure 2C).

In silico staining deep learning model

In silico IHC uses a trained neural network that takes an H&E-LFB-stained WSI as input and generates synthetic IHC-stained images predicting the presence of NFTs, A β plaques, and neuritic plaques (Figure 1). From the WSI, we select non-overlapping patches of 2,048 \times 2,048 pixels, corresponding with 517 μ m \times 517 μ m. Each patch is given to the trained neural network, which makes separate predictions for the probabilities of at least one amyloid plaque, NFT, or neuritic plaque appearing within the patch. The predicted probability for each patch in the WSI is mapped to colors imitating the real IHC, which are then

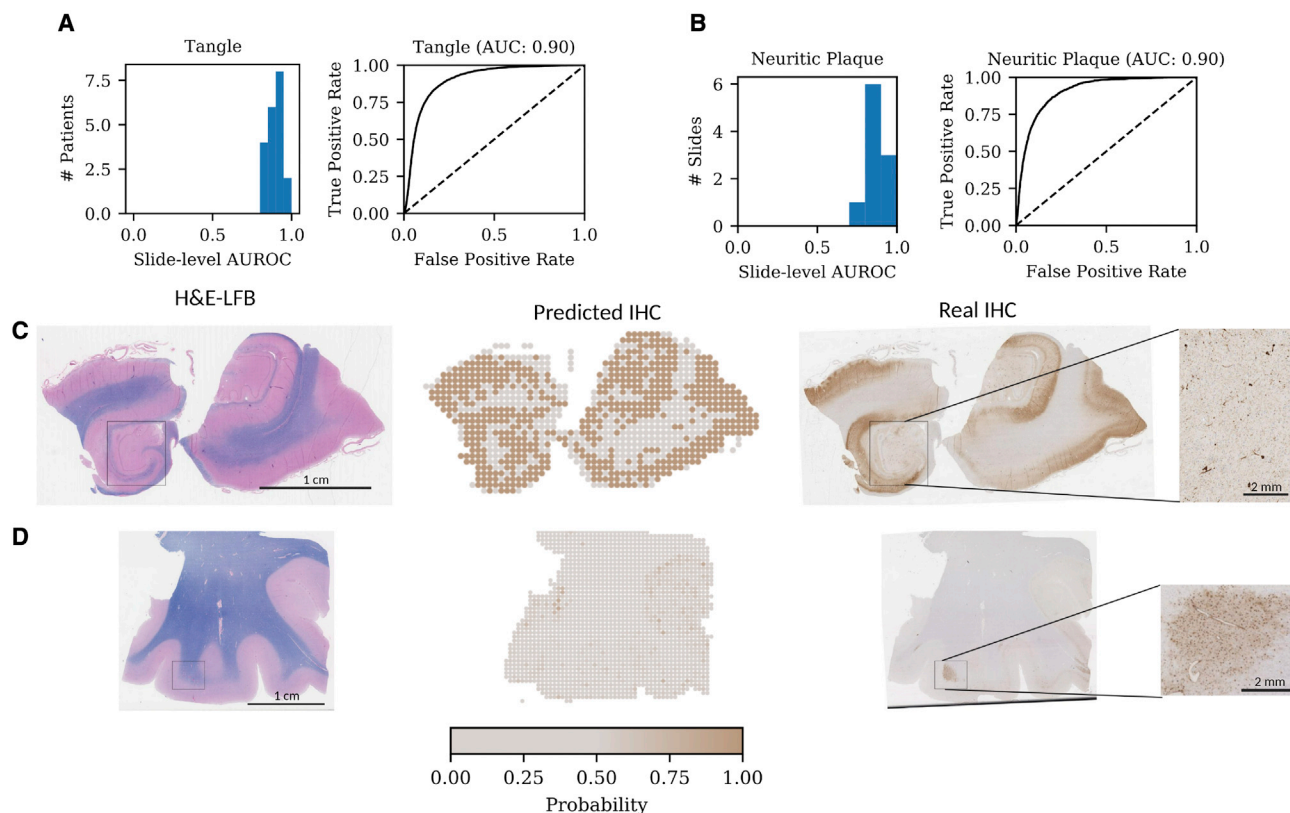


Figure 3. Results for NFT and neuritic plaque predictions

For visualization, the predictions from *in silico* IHC for NFTs and neuritic plaques are combined to match the real phospho-tau IHC, which will show positivity for both.

(A and B) AUROCs for individual slides and receiver operating characteristic curve for (A) NFTs and (B) neuritic plaques.

(C) Predictions for a highly stained slide. The model correctly identifies tangles in the gyrus, which are difficult to see on the real IHC at low resolution (inset).

(D) Predictions for a slide with low staining. The model correctly ignores Aging-related tau astrogliopathy (ARTAG), which include the Tau protein, but are not NFTs (inset).

combined into a synthetic image for each of the targets, which can be used to identify whether they are present, along with their locations.

Training *in silico* IHC consists of three main steps. First, we register the serial sections of each sample in the dataset so that each IHC WSI is aligned with the corresponding H&E-LFB image as much as possible. Second, we use the IHC WSIs to identify the patches containing NFTs, A β plaques, and neuritic plaques (examples in Figure S1). Third, we train a neural network to directly predict the presence of NFTs, A β plaques, and neuritic plaques from the H&E-LFB image patches (see the STAR Methods for more details).

To register the serial sections, we use the Oriented FAST and Rotated BRIEF feature detector (Ruble et al., 2011) to identify key points in each slide. Matching key points are selected using the RANSAC algorithm (Fischler and Bolles, 1981). The matched key points are then used to overlay the serial sections over each other. To assess the accuracy of the registration, we identified 50 pairs of blood vessels visible in both the H&E and one of the IHC slides. The registration process resulted in these blood vessels being mapped to an average of 224 pixels apart, with a SD of 156 pixels ($56 \pm 39 \mu\text{m}$). The average registration error is approx-

imately 10% the width of a patch, so pairs of H&E-LFB and IHC patches are closely related.

Next, to identify NFTs, A β plaques, and neuritic plaques, we annotated 500 phospho-tau and 500 A β IHC-stained patches for the presence of each. Using the annotated dataset, we train one annotator network to identify NFTs and neuritic plaques from the phospho-tau IHC slides and a second annotator network to identify A β plaques from the A β IHC slides. The annotator networks are then used to label each patch on the real IHC with the hallmark changes that are present.

Finally, we combine the registration and IHC quantification to train an end-to-end model for predicting the presence of NFTs, A β plaques, and neuritic plaques directly from the H&E-LFB image. To train this model, we use the registered slides to identify paired H&E-LFB patches and IHC patches. The annotator network uses the IHC slides to provide annotations of NFTs, A β plaques, and neuritic plaques, which are used as supervision to train *in silico* IHC.

Evaluation of *in silico* IHC

We run *in silico* IHC on a held-out test set of 83 samples (examples in Figures 3 and 4) from 19 patients to evaluate its ability to

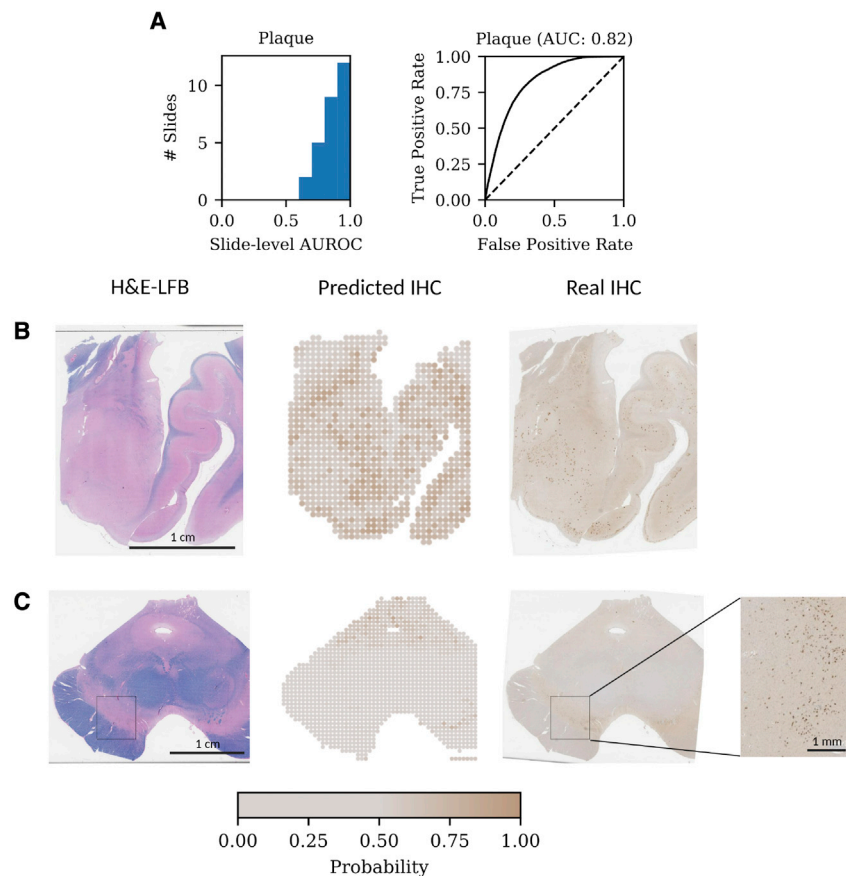


Figure 4. Results for amyloid plaque predictions

(A) AUROCs for individual slides and receiver operating characteristics curve.

(B) Prediction for a highly stained slide.

(C) Predictions for a negative slide. The model correctly ignores neuromelanin in the H&E-LFB slide, which are stained in the IHC but are not actual amyloid plaques (inset).

IHC, achieving AUROCs of 0.94 (95% CI, 0.92–0.95), 0.93 (95% CI, 0.91–0.94), and 0.83 (95% CI, 0.80–0.86) for NFT, neuritic plaques, and A β plaques, respectively (Figure S2).

The prevalence of neuropathological hallmark changes varies highly between different areas of the brain. For example, NFT, neuritic plaques and A β plaques are significantly more common in the hippocampus than the midbrain, and these lesions almost never appear in white matter. To evaluate the ability of *in silico* IHC to extract information beyond standard location features, we trained a logistic regression model to predict the appearance of hallmark changes using the region of the brain, the fraction of the patch stained with LFB (which identifies white matter), and the number of nuclei in the patch as features. This model achieves AUROCs of 0.78 (95% CI, 0.75–0.80) for NFTs, 0.80

(95% CI, 0.77–0.82) for neuritic plaques, and 0.78 (95% CI, 0.76–0.80) for A β plaques and is significantly outperformed by *in silico* IHC. This suggests that the computer vision algorithm can leverage more fine-grained morphological features of the tissue neighborhoods in its assessment.

identify regions with NFTs, neuritic plaques, and A β plaques. We find that *in silico* IHC archives AUROCs of 0.91 (95% CI, 0.88–0.95), 0.92 (95% CI, 0.87–0.94), and 0.88 (95% CI, 0.82–0.93), respectively. As additional verification that our results are not skewed by potential errors in the automated identification of lesions from IHC, we annotated the IHC patches corresponding to 250 random H&E patches in our test set as ground truth. When evaluated on these patches, *in silico* IHC achieves AUROCs of 0.92 (95% CI, 0.87–0.96), 0.90 (95% CI, 0.84–0.95), and 0.92 (95% CI, 0.84–0.97), respectively.

In the H&E-LFB-stained slides, there can be artifacts and other structures unrelated to AD disease, which could potentially be falsely identified as lesions. For example, there are commonly folds and tears in the tissue, and some regions of the brain contain pigmented regions such as neuromelanin and lipofuscin, which may seem to be abnormal. However, we find that *in silico* IHC can correctly identify that these regions are negative (Figure 4).

To further assess the reliability of our algorithm, we collected H&E-LFB and IHC images from 30 consecutive patients with neurodegenerative diseases. These images were taken at different times from the data used in the algorithm’s training and thus constitute a separate test set. We deployed the algorithm without any modification to this new test data. The *in silico* IHC matches well with the experimentally obtained

Additionally, we study the choice of neural network architecture used by *in silico* IHC by comparing against the performance of AlexNet (Krizhevsky et al., 2012), VGG-11 (Simonyan and Zisserman, 2015), and ResNet-18 (He et al., 2016). All models were trained using the same data. We find that *in silico* IHC can outperform the other choices of neural network architectures (Table S1).

Interpretation of *in silico* IHC predictions

To better understand the predictions made by our model, we use deep learning interpretation methods to provide attributions. We used the integrated gradients method (Sundararajan et al., 2017), which identifies pixels in an image that the model considers most useful for making a prediction. In Figure 5, we show several examples of attributions for neuritic and A β plaques. We additionally show the corresponding IHC stain and find that the attributions match the plaques in location and size, suggesting that *in silico* IHC has learned to identify individual lesions, despite only being trained on patch-level labels.

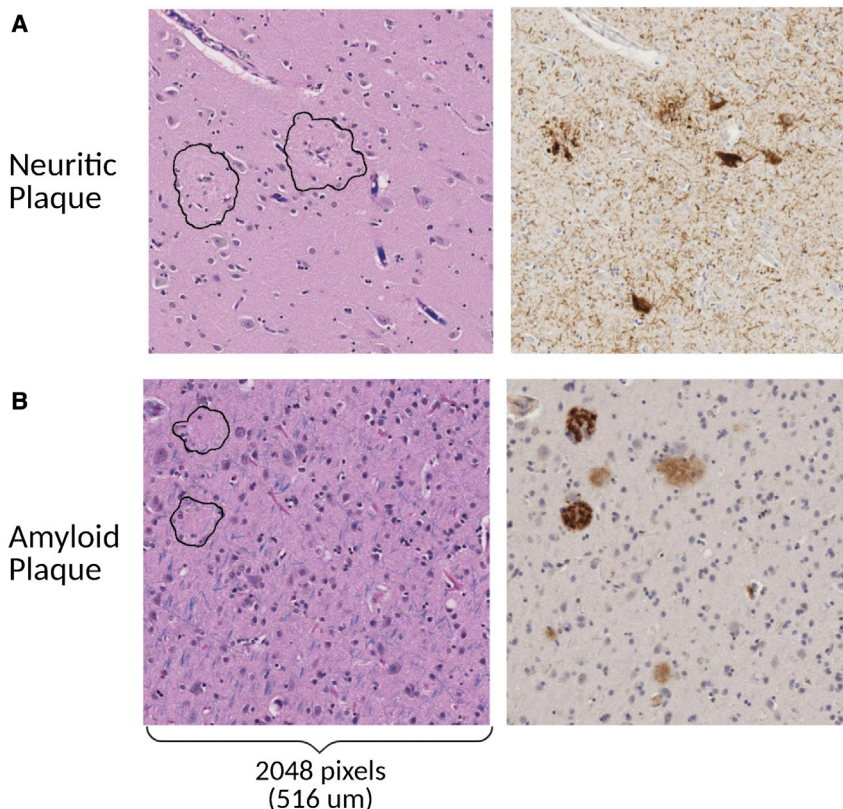


Figure 5. Attribution of *in silico* IHC predictions

(A and B) Raw H&E image, model attribution, and corresponding IHC for (A) phospho-Tau and (B) amyloid beta. The circled regions on the left are the regions in the image that the model pays the most attention in each H&E image. These regions have subtle differences in their morphology and texture that the computer vision model picks up. They match the areas where actual phospho-Tau and amyloid beta are located, as identified via experimental IHC (brown blobs in the right).

Neuritic Plaque

Amyloid Plaque

2048 pixels
(516 um)

learning models, decreasing the need for gathering expert annotations, allowing large datasets with fine-grained labels to be created. The methodology we propose for developing *in silico* IHC can be extended to these diseases and can aid in advancing future work in translation to other areas of medical imaging.

An exciting potential application of *in silico* IHC is to help pathologists quickly assess samples and prioritize samples for experimental immunostaining in both diagnostic and large cohort research settings. *In silico* staining also makes it easier for pathologists to visualize different biological information on the same tissue section compared to the

typical setting where one must mentally align stains taken on different sections.

DISCUSSION

In this work, we introduce a model for translating from standard H&E-LFB-stained neuropathological samples to synthetic phospho-tau and A β -stained images. Our model achieves high accuracy for classifying the presence of NFTs, neuritic plaques and A β plaques on independent test samples. Moreover, it significantly outperforms models using hand-crafted features based on information about the nuclei and region of the brain. We additionally use interpretation methods to identify regions that the model considered most relevant for making a classification and found that the regions closely match the areas identified by the real IHC, suggesting that the model has learned fine-grained morphological features of cellular neighborhoods that are indicative of the AD related plaques and tangles.

A complete analysis of a brain sample for neurodegenerative disease requires IHC-stained slides to be prepared for many regions of the brain, greatly increasing the cost and time needed for preparing the samples, limiting diagnostic workup outside of research settings, and severely limiting large-scale genomic-endophenotype association studies (Beecham et al., 2014; Flanagan et al., 2017). As a result, we focused on the most common neurodegenerative disease as a case study using *in silico* IHC for translating from routinely and cost-efficiently prepared H&E-LFB-stained slides to the necessary IHC stains.

In other tissues and diseases where immunostaining is used, paired H&E-LFB and IHC samples can be used to train deep

Limitations of the study

Our work is a proof-of-concept study that demonstrates the possibility of *in silico* IHC. More work is needed to harden this technique into a software that can be readily used in laboratories. It is also necessary to ensure variability in the histochemical slide preparation from other sources does not degrade the performance of *in silico* IHC. We note that *in silico* IHC is only able to identify pathologic changes that have a discernible disturbance in the H&E-LFB slides, and performance on pathologic change with less disruption may be more challenging to identify. For example, A β plaques result in less disruption than both NFTs and neuritic plaques, which may explain *in silico* IHC's weaker performance on A β plaques. It may also be more challenging to identify pretangles, which are precursors to NFTs, owing to their limited disruption on the H&E-LFB slides.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact

- Materials availability
- Data and code availability
- **METHOD DETAILS**
 - Data curation
 - Digital staining procedure
 - Registration of H&E-LFB and IHC slides
 - Identification of lesions from IHC
 - Training and evaluating the H&E-LFB model
 - Model architecture and training
 - Interpretation of *in silico*-IHC predictions
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.crmeth.2022.100191>.

ACKNOWLEDGMENTS

This work was supported by UF1 AG057707, NSF CAREER 1942926, NIH P30AG059307, 5RM1HG010023, RO1 AG021055, RO1 AG062695, UF1 AG057707, and grants from the Silicon Valley Foundation and the Chan Zuckerberg Initiative.

AUTHOR CONTRIBUTIONS

B.H., T.M., and J.Z. designed the experiments with inputs from all the authors. B.H. developed the algorithm and performed experiments. S.B., E.F., A.A., J.S., C.K., and M.C. provided materials and annotations. All the authors contributed to writing and reviewing the paper.

DECLARATION OF INTERESTS

The authors declare that there are no competing interests.

Received: August 5, 2021
Revised: December 23, 2021
Accepted: March 4, 2022
Published: March 28, 2022

REFERENCES

Beecham, G.W., Hamilton, K., Naj, A.C., Martin, E.R., Huentelman, M., Myers, A.J., Corneveaux, J.J., Hardy, J., Vonsattel, J.P., Younkin, S.G., et al. (2014). Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet.* *10*, e1004606.

Besser, L.M., Kukull, W.A., Teylan, M.A., Bigio, E.H., Cairns, N.J., Kofler, J.K., Montine, T.J., Schneider, J.A., and Nelson, P.T. (2018). The revised national alzheimer's coordinating center's neuropathology form-available data and new analyses. *J. Neuropathol. Exp. Neurol.* *77*, 717–726.

Campanella, G., Hanna, M.G., Geneslaw, L., Mirafior, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., and Fuchs, T.J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* *25*, 1301–1309.

Christiansen, E.M., Yang, S.J., Ando, D.M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'Neil, A., Shah, K., Lee, A.K., et al. (2018). In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* *173*, 792–803.e19.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Li, F. (2009). ImageNet: a large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 248–255.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* *546*, 115–118.

Fischler, M.A., and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* *24*, 381–395.

Flanagan, M.E., Marshall, D.A., Shofer, J.B., Montine, K.S., Nelson, P.T., Montine, T.J., and Keene, C.D. (2017). Performance of a condensed protocol that reduces effort and cost of NIA-AA guidelines for neuropathologic assessment of alzheimer disease. *J. Neuropathol. Exp. Neurol.* *76*, 39–43. <https://doi.org/10.1093/jnen/nlw104>.

Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., and Gerstung, M. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* *1*, 800–810.

He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundeberg, J., and Zou, J. (2020). Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* *4*, 827–834.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE), pp. 770–778.

Head, E., Corrada, M.M., Kahle-Wroblewski, K., Kim, R.C., Sarsoza, F., Goodus, M., and Kawas, C.H. (2009). Synaptic proteins, neuropathology and cognitive status in the oldest-old. *Neurobiol. Aging* *30*, 1125–1134.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE), pp. 4700–4708.

Hyman, B.T., Phelps, C.H., Beach, T.G., Bigio, E.H., Cairns, N.J., Carrillo, M.C., Dickson, D.W., Duyckaerts, C., Frosch, M.P., Masliah, E., et al. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's Dementia* *8*, 1–13.

Jackson, C.R., Sriharan, A., and Vaicukus, L.J. (2020). A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC and evaluation on primarily melanocytic neoplasms. *Mod. Pathol.* *33*, 1638–1648.

Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K., Bankhead, P., et al. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* *1*, 789–799.

Koga, S., Ghayal, N.B., and Dickson, D.W. (2021). Deep learning-based image classification in differentiating tufted astrocytes, astrocytic plaques, and neuritic plaques. *J. Neuropathol. Exp. Neurol.* *80*, 306–312.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks (Advances in neural information processing systems).

Levy-Jurgenson, A., Tekpli, X., Kristensen, V.N., and Yakhini, Z. (2020). Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci. Rep.* *10*, 18802.

Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., et al. (2018). 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience* *7*, gij065. <https://doi.org/10.1093/gigascience/gij065>.

Liu, S., Li, X., Zheng, A., Yang, F., Liu, Y., Guan, T., and He, Y. (2021). The generation of virtual immunohistochemical staining images based on an improved cycle-GAN. In Machine Learning and Intelligent Communications, pp. 137–147. https://doi.org/10.1007/978-3-030-66785-6_16.

Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* *5*, 555–570. <https://doi.org/10.1038/s41551-020-00682-w>.

Montine, T.J., Phelps, C.H., Beach, T.G., Bigio, E.H., Cairns, N.J., Dickson, D.W., Duyckaerts, C., Frosch, M.P., Masliah, E., Mirra, S.S., et al. (2012). National Institute on Aging-Alzheimer's Association guidelines for the

- neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* 123, 1–11.
- Montine, T.J., Monsell, S.E., Beach, T.G., Bigio, E.H., Bu, Y., Cairns, N.J., Frosch, M., Henrikson, J., Kofler, J., Kukull, W.A., et al. (2016). Multisite assessment of NIA-AA guidelines for the neuropathologic evaluation of Alzheimer's disease. *Alzheimer's Dementia* 12, 164–169.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9, 62–66. <https://doi.org/10.1109/tsmc.1979.4310076>.
- Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., and Zou, J.Y. (2020). Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 580, 252–256.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., and Desmaison, A. (2019). PyTorch: an imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32.
- Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydin, H., Zuckerman, J.E., Chong, T., Sisk, A.E., et al. (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* 3, 466–477.
- Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). ORB: an efficient alternative to SIFT or SURF. In 2011 International Conference on Computer Vision (ICCV) (IEEE), pp. 2564–2571.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>.
- Sharma, H., Zerbe, N., Klempert, I., Hellwich, O., and Hufnagl, P. (2017). Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput. Med. Imag. Graph.* 61, 2–13.
- Signaevsky, M., Prastawa, M., Farrell, K., Tabish, N., Baldwin, E., Han, N., Iida, M.A., Koll, J., Bryce, C., Purohit, D., et al. (2019). Artificial intelligence in neuropathology: deep learning-based assessment of tauopathy. *Lab. Invest.* 99, 1019–1029.
- Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.
- Sundararajan, M., Taly, A., and Yan, Q. (2017). Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, 70, pp. 3319–3328.
- Xu, J., Luo, X., Wang, G., Gilmore, H., and Madabhushi, A. (2016). A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* 191, 214–223.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
β-Amyloid	Biolegend	cat#800701
Phospho-Tau	ThermoScientific	cat#MN1020
Software and algorithms		
In silico-IHC	This paper	https://github.com/bryanhe/insilico-ihc (https://doi.org/10.5281/zenodo.6236002)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, James Zou (jamesz@stanford.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The histopathology images will be shared by the lead contact upon request.
- All original code has been deposited at Github (<https://github.com/bryanhe/insilico-ihc>) and Zenodo (<https://doi.org/10.5281/zenodo.6236002>).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Data curation

The training cohort comprised 160 cases drawn from the 90 + Study (Head et al., 2009). Tissues were sampled and analyzed as previously described (Montine et al., 2016; Besser et al., 2018). Regions analyzed for the training cohort included the substantia nigra at the level of the red nucleus, middle frontal gyrus or Brodmann area (BA)9, hippocampus at the level of the lateral geniculate nucleus, and amygdala. The additional validation cohort comprised 30 consecutive cases from the 90 + Study. For this validation cohort, each case consists of samples from the primary visual cortex (BA 17), substantia nigra at the level of the red nucleus, inferior parietal lobule (BA39), striatum at the level of the anterior commissure (caudate nucleus and putamen), and hippocampus at the level of the lateral geniculate nucleus. For evaluating AD neuropathologic change by consensus guidelines, Aβ staining on the inferior parietal lobule and middle frontal gyrus are the same score, Aβ staining on the striatum and amygdala are the same score, and phospho-tau staining on the primary visual cortex and middle frontal gyrus are the same score.

Tissues were stained histochemically with H&E-LFB, and immunohistochemically with antibodies to Aβ (4G8, Biolegend, cat#800701, working dilution 1:1,000) or phospho-Tau (AT8, ThermoScientific, cat#MN1020, working dilution 1:1,000). All slides are digitized at 40× magnification on a Leica AT2 scanner.

Digital staining procedure

In silico-IHC takes an H&E-LFB-stained WSI as input (Figure 1A). WSIs are typically around 100,000 × 100,000 pixels (>1 GB), which are too large to process with a neural network. To handle this large size, *in silico*-IHC first divides the WSI into 2,048 × 2,048 patches (517 μm × 517 μm). Each patch is then passed through our trained neural network, resulting in a separate prediction for the probabilities that the patch contains an Aβ plaque, NFT or neuritic plaque. The predictions for the patches are then merged as a synthetic IHC stained image by representing each patch with a colored spot based on the probability of containing each lesion.

Registration of H&E-LFB and IHC slides

To train our system, we use serial sections of tissue stained with H&E-LFB, phospho-tau IHC, and Aβ IHC (Figure 1B). The samples are collected from five different regions: amygdala, hippocampus, contralateral hippocampus, midbrain, and BA9. The sections from

a sample are closely related owing to their serial nature (5 μm between sections), but cutting the sample results in the exact spatial relationship between the sections being destroyed: slight distortions in the tissue will result from the cutting process, and the images will be translated and rotated owing to the sections not being in the same position on the slide when digitizing (Figure S3). The first step in our training procedure is then to register the IHC slides to the H&E-LFB slides to provide paired examples for the neural network.

To avoid this issue, we begin by using Otsu's method to threshold the H&E-LFB and IHC slides into foreground and background (Otsu, 1979). After binarizing the images, the different color schemes of the stains are no longer an issue, but the main features of the sample are still visible.

Next, we identify candidate key points using the ORB feature detector. The ORB feature detector identifies areas of the image with distinctive structures (e.g., sharp corners of the tissue) as candidate key points. For each key point, the ORB feature detector provides a descriptor, a real-valued vector, that allows the key point to be matched across images.

With the candidate key points and descriptors for a pair of H&E-LFB and IHC slides, we create a matching between the key points from the two slides with the most similar descriptors. This process results in many correct matches, but will also include incorrect matches that must be filtered before finding a transform between the images. We use the RANSAC method for identifying the outliers from the matches, and we fit an affine transform on the remaining matches to register the images (Figure S3). We run RANSAC for 2,000 iterations and consider a key point as an inlier if the error is within 25 pixels.

Identification of lesions from IHC

The next step in generating the expected output for training the neural network is identifying the hallmark lesions from the IHC images. The IHCs can have considerable variation in the background color along with other unrelated structures such as folds in the tissue, lipofuscin, and neuromelanin, which must be distinguished from the NFTs and A β plaques. To handle these issues, we selected 500 A β and 500 phospho-tau IHC regions of 16,384 \times 16,384 pixels (4,161 μm \times 4,161 μm) from the WSIs in the training set. The A β patches were annotated for instances of A β plaques, and the phospho-tau patches were annotated for instances of NFTs and neuritic plaques. The regions were then divided into 2,048 \times 2,048-pixel patches, which were considered positive for each lesion if any instance of the lesion appeared within the patch.

With these annotated patches, we trained two separate DenseNet121 models (Huang et al., 2017) to identify A β plaques from the A β IHC slides and identify NFTs and neuritic plaques from the phospho-tau IHC slides using 2,048 \times 2,048-pixel patches extracted from the larger patches. Our models were implemented using the PyTorch library (Paszke et al., 2019). Our trained model achieved AUCs of 0.98 for NFTs, 0.99 for neuritic plaques, and 0.97 for A β plaques (Figure S4).

Training and evaluating the H&E-LFB model

With the combined results of registration and identification from IHC, we have patch-level annotations for the H&E-LFB slides. For each H&E-LFB patch, we identify the corresponding IHC patch and run our trained IHC model on the patch to identify if A β plaques, NFTs, or neuritic plaques are present. From our training set, we extract 190,992 patches, and train a Densenet121 model to predict both the presence of A β plaques, NFTs, and neuritic plaques. We evaluate the performance of the model on the patches from the held-out test patients corresponding to IHC patches that were confidently classified for each hallmark change (<5% or >95%). We find that the model's predicted probabilities of the hallmark changes are closely aligned to the true fraction of positive patches (Figure S5).

Model architecture and training

For generating predictions with *in silico*-IHC, we use a DenseNet121 architecture (Huang et al., 2017), which has previously been shown to perform well on the ImageNet dataset (Deng et al., 2009; Russakovsky et al., 2015). The DenseNet121 architecture consists of 120 convolutional layers arranged into 4 densely connected blocks, followed by a fully connected layer. We initialize the model with pretrained ImageNet weights, and we fine-tune all parameters in the model for 150 epochs. We use a stochastic gradient descent optimizer with an initial learning rate of 1×10^{-4} and a momentum of 0.9, and we decay the learning rate by a factor of 10 every 50 epochs. The optimizer trains the model by minimizing the binary cross-entropy loss between the model's predictions and the label for each hallmark change extracted from the matched IHC patch. During training, we augment the dataset by including all rotations and reflections of the patches. For our final evaluation, we select the model from the epoch with the highest AUC on the validation set.

Interpretation of *in silico*-IHC predictions

To interpret the predictions made by *in silico*-IHC, we first use the integrated gradients method to provide per-pixel attributions for each patch. We then sought to identify regions, rather than pixels, that resulted in positive predictions. First, we applied a Gaussian filter to the attributions with a standard deviation of 5 for the Gaussian kernel; this allows nearby regions with high attributions to be connected. Next, we identified pixels with attributions above the 90th percentile, and extracted connected regions of pixels. Regions smaller than 500 pixels (32 μm^2) were then filtered out. Finally, the contours of the remaining regions were then extracted.

QUANTIFICATION AND STATISTICAL ANALYSIS

CIs for the AUROCs in the results were computed using 10,000 bootstrapped samples and obtaining 95 percentile ranges for each prediction. The performance of *in silico*-IHC for NFT and neuritic plaque predictions were computed for 51 samples with phospho-tau staining in the test set, and the performance for amyloid plaque predictions were computed for 53 samples with A β staining in the test set. The performance of *in silico*-IHC on the additional evaluation data was computed for 60 samples with phospho-tau staining and 90 samples with A β staining. The performance of the identification of lesions from IHC was computed using 75 labeled patches in the test set. Additional analysis details are provided in the Results section and in figure legends. Statistical analysis was performed using Python.