Deep Learning Algorithms for Corneal Amyloid Deposition Quantitation in Familial Amyloidosis

Klaus Kessel^a Jaakko Mattila^b Nina Linder^{a, c} Tero Kivelä^{d, e} Johan Lundin^{a, f}

^aInstitute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; ^bCornea Service, Department of Ophthalmology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; ^cDepartment of Women's and Children's Health, International Maternal and Child Health (IMCH), Uppsala University, Uppsala, Sweden; ^dOphthalmic Pathology Laboratory, Department of Ophthalmology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; ^eOphthalmic Pathology, Hospital District of Helsinki and Uusimaa Laboratory (HUSLAB), Helsinki, Finland; ^fDepartment of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

Keywords

Familial amyloidosis, Finnish · Meretoja syndrome · Corneal amyloidosis · Gelsolin · Machine learning

Abstract

Objectives: The aim of this study was to train and validate deep learning algorithms to quantitate relative amyloid deposition (RAD; mean amyloid deposited area per stromal area) in corneal sections from patients with familial amyloidosis, Finnish (FAF), and assess its relationship with visual acuity. **Methods:** Corneal specimens were obtained from 42 patients undergoing penetrating keratoplasty, stained with Congo red, and digitally scanned. Areas of amyloid deposits and areas of stromal tissue were labeled on a pixel level for training and validation. The algorithms were used to quantify RAD in each cornea, and the association of RAD with visual acuity was assessed. **Results:** In the validation of the amyloid area classification, sensitivity was 86%, specificity 92%, and F-score 81. For corneal stromal area classification, sensitivity was 74%, specificity 82%, and F-score 73. There was

insufficient evidence to demonstrate correlation (Spearman's rank correlation, -0.264 , $p = 0.091$) between RAD and visual acuity (logMAR). **Conclusions:** Deep learning algorithms can achieve a high sensitivity and specificity in pixel-level classification of amyloid and corneal stromal area. Further modeling and development of algorithms to assess earlier stages of deposition from clinical images is necessary to better assess the correlation between amyloid deposition and visual acuity. The method might be applied to corneal dystrophies as well.

© 2019 S. Karger AG, Basel

Introduction

Corneal amyloid deposition, formerly incorrectly known as lattice corneal dystrophy, type II [1], is the earliest clinical finding in dominantly inherited familial amyloidosis, Finnish (FAF; OMIM 105120), also known as Meretoja syndrome and hereditary gelsolin amyloidosis [2–4]. The corneal changes consist of lace-like amy-

loid deposits (lattice lines) that are most prominent in the peripheral cornea, corneal erosions [2, 5], and superficial stromal scarring [6]. FAF results from a single base mutation in the gelsolin (GSN) gene on chromosome 9q33 that replaces guanine with either adenine [7, 8] or thymidine [9] and results in coding of aspartic acid or tyrosine, respectively, instead of asparagine. The abnormal degradation product of gelsolin is additionally deposited in other ocular tissues [10] and throughout the body [2, 3].

Congo red [11, 12] and immunohistochemistry [12] have helped to localize lattice lines predominantly in the anterior and mid stroma. A nearly continuous layer of amyloid commonly accumulates under Bowman's layer that is in advanced cases discontinuous or scarred, and a thinner layer may be present at the epithelial basement membrane [12]. Similar findings are evident with confocal microscopy [5, 6]. Dense opaque deposits are seen posterior to pleomorphic basal epithelial cells. Other findings include irregular and thickened Bowman's layer and stromal filaments that correspond to lattice lines [5, 6]. The subepithelial deposits, recurrent erosions, and stromal scarring, rather than the lattice lines, cause visual loss. Keratoplasty mostly improves vision temporarily because of a high risk of graft failure [13].

The relationship between amyloid deposition and visual acuity has not been previously quantitated. We trained and applied deep learning algorithms to quantify the relative area of amyloid in corneal sections from patients with FAF to evaluate its association with vision. Deep neural networks in image analysis learn features in training images and, once trained, can be used to classify new images or areas of them [14]. Such networks have been applied to a large variety of image-based classification and quantification tasks, e.g., in cancer and infectious disease diagnostics [15, 16] and grading and diagnosis of retinopathy [17–19]. Our method can be used to calculate relative amyloid deposition (RAD) in other diseases as well.

Materials and Methods

Patients

Tissue samples were obtained from 42 patients with FAF who had undergone penetrating keratoplasty at the Department of Ophthalmology, Helsinki University Hospital, Finland, between 1990 and 2010 to improve vision. The cornea was fixed in formalin and embedded in paraffin, and 5- μ m-thick sections were cut and stained with Congo red. The samples were pseudonymized by as-

signing a numeric code. Retrospective visual acuity data from patient charts were obtained by consent of the review board of the Helsinki and Uusimaa Hospital District, and the study followed the tenets of the Declaration of Helsinki.

Slide Preparation and Digitalization

Each microscopic slide was prepared to contain 4–12 corneal sections from a single patient. The slides were digitized using a slide scanner (Pannoramic 250 Flash, 3DHISTECH Ltd., Budapest, Hungary), with $\times 20$ objective magnification and $\times 1.6$ camera adapter magnification. One pixel in the whole slide image (WSI) corresponds to a $\sim 0.24 \times 0.24 \mu\text{m}$ area in the sample. The images were stored on a WSI management server (WebMicroscope, Aiforia Technologies Oy, Helsinki, Finland) where they were organized according to their pseudonymization code.

Labeling and Training

First, we drew a large bounding box around most of a corneal section on a WSI (Fig. 1a). We then used an image splitter tool (WebMicroscope, Aiforia Technologies Oy) to create 300×300 pixel nonoverlapping tiles ($n = 545$) of the selected area. These were used for initial training and validation of the algorithms. Prior to further training, by consensus of 2 ophthalmologists (J.M. and T.T.K.), 4 histologic regions were identified that are distinct and, therefore, should be represented equally in training: amyloid surrounded by stroma or immediately under Bowman's layer, epithelium with or without amyloid, endothelium with or without amyloid, and stroma with no amyloid deposits. The histologic region classes were selected to represent different types of background for amyloid deposits, i.e., morphological entities within which amyloid can occur. Using the least distorted corneal section in each WSI (Fig. 1b), we captured equal numbers of 909×909 pixel image tiles ($n = 900$) from each of these regions in the WSI (Fig. 1c–f).

We labeled the corneal stroma, including its collagen lamellae, fibroblasts, Bowman's layer, and separately amyloid on a pixel level in tiles. Empty space between lamellae was left out. We used an image-level label to mark slides with no stroma. Amyloid areas that were best demarcated and well contrasted relative to surrounding tissue were similarly labeled. Amyloid-free images were labeled as such with an image-level label.

Two classifier algorithms were created, one for detection of amyloid deposited area and another for detection of corneal stromal area (WebMicroscope 3.0, Aiforia Technologies Oy). For training of the amyloid detection algorithm, we used color input tiles (RGB), set the feature size or size of the area to be analyzed at any one point during training to 40 pixels (9.6 μm), and epochs or number of forward and backward passes or analyses of the data to 50. For the corneal stroma algorithm, we set feature size to 300. A supervised mode was used for the training of both algorithms, where the algorithm learns to associate features of labeled areas with the respective label. Half of the labeled tiles were randomly drawn for training (training set) and the other half for validation of the algorithms (validation set).

The trained algorithms predicted whether areas of tiles they processed resembled the target label or not. Areas predicted with a high degree of certainty (score) were classified with the target label. A ROC curve was generated for each algorithm (online suppl. Fig. 1; for all online suppl. material, see www.karger.com/doi/10.1159/000500896; amyloid algorithm ROC [left] and cor-

Fig. 1. Algorithm training. Corneal tissue sections stained with Congo red (**a, b**) used for training. Four types of histologically distinct regions were captured from corneal sections as image tiles: stromal/Bowman's layer amyloid deposits (**c**), epithelium (**d**), endothelium (**e**), and stroma with no or minimal amyloid deposits (**f**). Amyloid deposits appear in a distinctive red color (arrow in tiles **c** and **d**). **a** Scale bar = 400 μm . **c** Scale bar = 50 μm .

neal stroma algorithm ROC [right] in the validation images). The cutoff/threshold score for detection was optimized to minimize false positives.

We performed a statistical analysis of the validation image set to assess the performance of algorithms. We also visually assessed the algorithms by applying them to a so-called "test set" containing $2,000 \times 2,000$ pixel unlabeled tiles from corneal sections distinct from those used in training and validation. To further decrease the number of false negatives and false positives, we manually added some tiles with artifacts, unusual morphology, tissue fragments, and folds in the corneal sections to the training set to ensure that such features would be understood by the algorithm. These added tiles were of sizes ranging from $2,000 \times 2,000$ to 300×300 pixels.

The entire set of tiles collected for training and validation purposes consisted of 1,785 image tiles. Of these tiles, 472 were used for training the corneal stroma algorithm, and 176 were used for training the amyloid algorithm.

Application of the Algorithms

Each corneal section area in WSIs was bounded by a box to select that area for analysis. These areas were divided into smaller, more manageable nonoverlapping tiles and then analyzed with the trained algorithms using in-house middleware (FIMM Image Analyzer, Institute for Molecular Medicine Finland, Helsinki, Finland). Areas classified by the algorithms as either amyloid or stroma were separately presented with a red overlay. Each of the analyzed WSIs were subsequently uploaded back to the WSI management system, where they can be viewed. The areas detected were quantified as pixels.

Statistical Methods

Sensitivity, positive predictive values (PPVs) and F-scores (weighted average of recall and precision) for each classifier were also calculated on a pixel level in the validation set.

To assess association between RAD and visual acuity, we first checked for normality of data using the Shapiro-Wilk test. Then, we used Spearman's rank correlation coefficient.

Results

In untrained, labeled tiles (validation set), the sensitivity for detection of amyloid labeled regions was 86%, with a specificity of 92%, a PPV of 76%, and an F-score of 81. The corresponding sensitivity for detection of stroma regions was 74%, with a specificity of 82%, a PPV of 72%, and an F-score of 73.

Visual Assessment of Algorithm Classification

To visually assess the algorithms, we compared classified tiles with unclassified tiles in the test set. Regions classified as amyloid or stroma were visualized with a bright red overlay. Well-demarcated and contrasted amyloid deposits in the stroma, Bowman's layer, and near to Descemet's membrane were detected (Fig. 2a, label). Typical false-positive areas included darkly stained la-

Fig. 2. Algorithm testing. Corneal whole slide image tiles stained with Congo red used for testing the amyloid algorithm (**a**, **b**) and the corneal stroma algorithm (**c**, **d**). Areas enclosed by dotted red lines are those detected by the respective algorithm (**a–d** label). Scale bar = 50 μ m.

Fig. 3. Algorithm application (analysis of specimen #37748). Corneal whole slide image stained with Congo red. Algorithm detection of amyloid (**a** label) and of corneal stroma (**b** label); detected areas are shown with a red overlay. Scale bar = 145 μ m.

mellae and tissue fragments as well as small areas of darkly stained endothelium, especially at folds in the section (Fig. 2a, label: arrow). False-negative amyloid classification typically occurred in cases where amyloid staining was more homogenous with surrounding tissues (Fig. 2b, label). In these cases, amyloid was more darkly or more lightly stained than typical training examples.

The corneal stroma classifier algorithm detected Bowman's layer, stromal lamellae, amyloid, and fibroblasts. Basal epithelial cells and Descemet's membrane were on the demarcating line. Typical false-positive classification areas included cellular clusters or tissue fragments resembling corneal tissue on the slide outside of the corneal section (Fig. 2c, label: arrow). In some instances, the epithelial area was falsely classified as stromal tissue. Typical false-negative areas included folds in the section (Fig. 2d, label: arrow).

Patient Analysis

All corneal sections from each patient (entire WSI) were analyzed. The sum of amyloid deposited area and the sum of corneal stroma area detected by the respective algorithms were collected per patient. Detected areas of amyloid (Fig. 3a, label) and of corneal stroma (Fig. 3b, label) were shown as a red, semi-transparent overlay. Detected areas from each patient can be viewed from the WSI management system (<http://fimm.webmicroscope.net/Research/faf>).

The mean amyloid deposited area per section was calculated by dividing the total amyloid area detected from all sections in a WSI by the number of sections on the WSI. The mean corneal stromal area per section was similarly calculated. The ratio of mean amyloid area to mean stromal tissue area was calculated and expressed as a percentage for each WSI to estimate the RAD in each cornea. The RAD for each patient is shown in Table 1.

RAD values ranged from 0.22 to 6.4%, and the median value was 2.42%. The association between RAD and visual acuity is shown in Figure 4. The Shapiro-Wilk test yielded a p value of 0.020 for RAD and 0.87 for visual acuity measured on the logarithm of the minimum angle of resolution (logMAR) scale. Thus, we can conclude that RAD is not normally distributed. Spearman's rank correlation was -0.2640 with $p = 0.091$. Thus, we have insufficient statistical evidence in our pilot series to conclude that a correlation exists between corneal amyloid deposition and visual acuity.

Discussion

Training deep learning algorithms to detect amyloid and corneal stroma involved labeling these areas in scanned corneal slides on a pixel level. The algorithms were then applied to WSIs of keratoplasty specimens where they detected areas of amyloid and corneal stroma. We were able to estimate the RAD in each cornea to

Table 1. WSI ID, RAD, and the corresponding patient's visual acuity

Slide ID	RAD	Visual acuity, logMAR
37731	1.98%	2.2
37732	4.31%	1
37733	2.54%	1.3
37734	1.90%	1
37735	2.42%	1.3
37736	5.19%	1.6
37737	2.57%	0.5
37738	2.49%	1
37739	3.78%	1.1
37746	2.60%	1.9
37740	0.99%	1.9
37745	3.50%	0.7
37747	2.61%	1.6
37741	2.50%	1.3
37773	2.81%	1.6
37748	2.26%	0.7
37742	1.83%	0.4
37753	2.71%	1.1
37749	3.27%	1.3
37750	1.14%	1.6
37756	0.64%	1.3
37755	0.22%	1.6
37743	1.07%	1.3
37758	1.27%	1
37744	0.84%	1.6
37730	6.40%	1
37759	1.21%	1.6
37757	1.76%	1.9
37752	5.68%	0.7
37762	2.60%	0.7
37760	1.85%	1.6
37761	2.74%	0.4
37751	3.52%	0.7
37754	1.75%	1
37768	2.22%	1.2
37774	0.43%	1.2
37767	1.45%	1.3
37770	3.13%	1.6
37764	4.42%	1.6
37766	1.55%	1.2
37771	1.63%	1.3
37772	1.41%	1

WSI, whole slide image; RAD, relative amyloid deposition.

attempt statistical analysis of its association with visual acuity.

Limitations of our study include the accuracy of measurement of algorithms trained by labeling on a pixel level. The quality of pattern detection is greatly

dependent on the quality of the labels. Amyloid labeled with Congo red that was well contrasted and well demarcated was a relatively easy pattern to detect using a deep neural network, trained with few examples. However, labeling of amyloid deposits in many cases proved difficult as they were not always well demarcated, and many were too small, or the stain was diffuse in pattern to digitally label them accurately. It is, however, also possible that some of the weak diffuse staining was background staining. We attempted to solve this problem by only labeling tiles with well-contrasted and well-demarcated amyloid deposits when training the network. This limited the total number of labels in the training and validation set but, on the other hand, allowed the annotator to assign labels with higher confidence. Ideally, the method would use amyloid detection by polarized light, but birefringence of collagen then becomes a problem. The importance of pixel-level labels with minimum false positives and negatives has been discussed in other studies [20, 21]. As labeled areas are not pixel-level precise and they are to some extent subjective, the accuracy of the generated classifier algorithms suffers from this method.

The subepithelial amyloid deposits lead to epithelial erosions and, thus, scarring of the corneal stroma. This process is the suspected cause of loss of vision in FAF patients. The lack of a clear correlation between corneal amyloid and visual acuity may be because operated patients, unlike other patients, had advanced corneal disease where increases in corneal amyloid may no longer alter visual acuity. Also, amyloid in lattice lines has been thought to be relatively unimportant for vision loss in FAF. Using cases with better visual acuity, counting only the deposits in the superficial corneal stroma, or both, could be used to test these hypotheses. The number of samples, limited by the rarity of FAF, may also have been insufficient to demonstrate a correlation.

To our knowledge, deep neural networks have not previously been applied to corneal specimens obtained at surgery. Other applications of deep learning in medical pathology and diagnostics have been explored using image-level labels. For example, diabetic retinopathy classification has been performed by assigning a retinopathy label to a fundus image and allowing the classifier algorithm to associate features with this desired output [17]. Classification of dermatological lesions has also been explored using this technique [16].

Measurement of the amyloid proportion retrospectively helps the clinician to better ascertain the stage of the disease. In future, this may be applied not only to patients

Fig. 4. Scatter plot of RAD and visual acuity (logMAR). RAD, relative amyloid deposition.

with FAF but also to those with various corneal dystrophies after penetrating keratoplasty. With further development, it could possibly also be applied to tissue imaging using confocal microscopy or anterior segment optical coherence tomography, and so might be used in clinical staging and follow-up in earlier stages of the disease and when contemplating surgical interventions.

Acknowledgements

We thank the founder of the project, late professor Juha Holopainen, and the corneal surgeons who operated the patients over several decades. We also thank Hakan Kucukel for developing the platform for applying trained algorithms to corneal samples, viewing results, and exportation of data. We also thank the FIMM Dig-

ital Microscopy and Molecular Pathology core facility and HiLIFE research infrastructure supported by University of Helsinki and Biocenter Finland.

Statement of Ethics

Retrospective visual acuity data from patient charts were obtained by consent of the review board of the Helsinki and Uusimaa Hospital District, and the study followed the tenets of the Declaration of Helsinki.

Disclosure Statement

Johan Lundin is co-founder and consultant for Aiforia Technologies Oy. The authors have no additional conflicts of interest.

Funding Sources

The study was funded by Evald and Hilda Nissi Foundation, The Finnish Eye and Tissue Bank Foundation, Finska Läkaresällskapet, Medicinska understödsföreningen Liv och Hälsa, Sigrd Jusélius stiftelse.

Author Contributions

T.T.K. and J.M. conceived the idea of studying amyloid concentration, and J.L. and N.L. the use of deep learning to achieve this. T.T.K. and J.M. collected the slides for analysis and sought ethical approval for the project. J.L. and N.L. designed the computational experiments for algorithm training. J.L. and K.K. planned the annotation and training of data, and T.T.K. and J.M. provided guidance with labeling. K.K. trained and validated the algorithms under the supervision of J.L.; K.K. then implemented them on collected samples. J.L. and T.T.K. provided critical revision of the manuscript. All aforementioned authors were involved in writing the manuscript and approved the final version to be published.

References

- Weiss JS, Möller HU, Aldave AJ, Seitz B, Bredrup C, Kivelä T, et al. IC3D classification of corneal dystrophies—edition 2. *Cornea*. 2015 Feb;34(2):117–59.
- Meretoja J. Familial systemic paramyloidosis with lattice dystrophy of the cornea, progressive cranial neuropathy, skin changes and various internal symptoms. A previously unrecognized heritable syndrome. *Ann Clin Res*. 1969 Dec;1(4):314–24.
- Kiuru-Enari S, Haltia M. Hereditary gelsolin amyloidosis. *Handb Clin Neurol*. 2013;115: 659–681.
- Nikoskinen T, Schmidt EK, Strbian D, Kiuru-Enari S, Atula S. Natural course of Finnish gelsolin amyloidosis. *Ann Med*. 2015;47(6): 506–11.
- Rothstein A, Auran JD, Wittpenn JR, Koester CJ, Florakis GJ. Confocal microscopy in Meretoja syndrome. *Cornea*. 2002 May;21(4):364–7.
- Rosenberg ME, Tervo TM, Gallar J, Acosta MC, Müller LJ, Moilanen JA, et al. Corneal morphology and sensitivity in lattice dystrophy type II (familial amyloidosis, Finnish type). *Invest Ophthalmol Vis Sci*. 2001 Mar; 42(3):634–41.
- Maury CP, Kere J, Tolvanen R, de la Chapelle A. Finnish hereditary amyloidosis is caused by a single nucleotide substitution in the gelsolin gene. *FEBS Lett*. 1990 Dec;276(1-2):75–7.
- Levy E, Haltia M, Fernandez-Madrid I, Kivunnen O, Ghiso J, Prelli F, et al. Mutation in gelsolin gene in Finnish hereditary amyloidosis. *J Exp Med*. 1990 Dec;172(6):1865–7.
- de la Chapelle A, Tolvanen R, Boysen G, Santavy J, Bleeker-Wagemakers L, Maury CP, et al. Gelsolin-derived familial amyloidosis caused by asparagine or tyrosine substitution for aspartic acid at residue 187. *Nat Genet*. 1992 Oct;2(2):157–60.
- Kivelä T, Tarkkanen A, Frangione B, Ghiso J, Haltia M. Ocular amyloid deposition in familial amyloidosis, Finnish: an analysis of native and variant gelsolin in Meretoja's syndrome. *Invest Ophthalmol Vis Sci*. 1994 Sep;35(10):3759–69.

- 11 Meretoja J. Comparative histopathological and clinical findings in eyes with lattice corneal dystrophy of two different types. *Ophthalmologica*. 1972;165(1):15–37.
- 12 Kivelä T, Tarkkanen A, McLean I, Ghiso J, Frangione B, Haltia M. Immunohistochemical analysis of lattice corneal dystrophies types I and II. *Br J Ophthalmol*. 1993;77:799–804.
- 13 Mattila JS, Krootila K, Kivelä T, Holopainen JM. Penetrating keratoplasty for corneal amyloidosis in familial amyloidosis, Finnish type. *Ophthalmology*. 2015 Mar;122(3):457–63.
- 14 Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. *Neurocomputing*. 2016; 187:27–48.
- 15 Bychkov D, Turkki R, Haglund C, Linder N, Lundin J. Deep learning for tissue microarray image-based outcome prediction in patients with colorectal cancer. *Proc SPIE*. 2016;9791: 979115.
- 16 Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb;542(7639):115–8.
- 17 Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec; 316(22):2402–10.
- 18 Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017 Dec; 318(22):2211–23.
- 19 Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RV, et al.; Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. *JAMA Ophthalmol*. 2018 Jul;136(7):803–10.
- 20 Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016 Jul;7(1):29.
- 21 Cruz-Roa A, Basavanthally A, González F, Gilmore H, Feldman M, Ganesan S, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. *SPIE Med Imaging*. 2014 March;904103.