Towards Automated Eye Movement Characterization for Stroke Patients Using Synthetic Video Data and Machine Learning

Citation for the original published paper (version of record):
Jalo, H., Ström, E., Ollila, S. et al (2025). Towards Automated Eye Movement Characterization for Stroke Patients Using Synthetic Video Data
and Machine Learning. Studies in health technology and informatics, 329: 951-955.
http://dx.doi.org/10.3233/SHTI250980

(article starts on next page)

MEDINFO 2025 — Healthcare Smart × Medicine Deep
M.S. Househ et al. (Eds.)

doi:10.3233/SHTI250980

Towards Automated Eye Movement Characterization for Stroke Patients Using Synthetic Video Data and Machine Learning

Hoor JALO[a,1], Eddie STRÖM[a], Samuel OLLILA[a], Robin KHATIRI[a], Jacob WESTERBERG[a], Teodor SVENSSON[a], Petra REDFORS[b,c], Katarina JOOD[b,c], Bengt Arne SJÖQVIST[a] and Stefan CANDEFJORD[a]

[a] *Chalmers University of Technology, Gothenburg, Sweden*
[b] *University of Gothenburg, Gothenburg, Sweden*
[c] *Sahlgrenska University Hospital, Gothenburg, Sweden*

Abstract. Stroke is a critical medical emergency that can cause permanent disability or death. Rapid identification of stroke, especially in prehospital settings, is crucial for timely treatment. Video analysis and machine learning (ML) could facilitate the prehospital assessment of stroke, but a lack of video data from stroke patients remain a barrier to developing effective models. This study explores the use of synthetic data to develop ML models, generating 73 videos mimicking characteristic eye movements of stroke patients through 3D modeling and animation. Four ML models were developed. Long short-term memory (LSTM) and gated recurrent units (GRU) achieved the best performance (over 84% in accuracy, precision, sensitivity, specificity and F1-Score). These findings highlight the promise of synthetic data for developing ML models for healthcare applications and the potential of ML-driven video analysis in the automated assessment of stroke-related eye movements, supporting advancements in prehospital stroke care.

Keywords. Stroke, video analysis, machine learning, synthetic data, eye movement

1. Introduction

Stroke is a critical medical emergency and a leading cause of mortality and disability worldwide [1]. Stroke is caused by either bleeding in the brain, called hemorrhagic stroke, or a clot obstructing arterial blood flow called ischemic stroke [2]. Ischemic stroke is best treated with thrombolysis (clot dissolving drug) followed by thrombectomy (mechanically removing clot) to restore blood flow in the brain and improve patient outcomes [2]. Stroke is assessed using various clinical stroke scales s such as the National Institute of Health Stroke Scale (NIHSS), a comprehensive and clinically validated scale that can measure severity of stroke with moderate to good accuracy [3,4]. It is however not practical in prehospital settings due to its complexity and demand on experience [3,5]. Simpler clinical scales have thus been introduced such as Face Arm Speech Test (FAST) and modified NIHSS (mNIHSS) to be used in the prehospital settings [5]. The American

[1] Hoor Jalo: hoor@chalmers.se. Hörsalsvägen 11, 412 96 Gothenburg, Sweden.

Heart Association (AHA) guidelines emphasizes the need for enhanced tools to characterize stroke in prehospital settings and calls for further research into bypass algorithms that help triage stroke patients to the most suitable treatment centers [5].

Initiatives are already underway to facilitate and improve stroke prehospital assessment, e.g. by digitalization [2,5,6]. In the ViPHS project (Video Support in the Prehospital Stroke Chain) video streaming from three cameras in ambulances is used to facilitate NIHSS evaluation by experienced stroke neurologist at distance [2]. ViPHS has been evaluated in realistic simulations, and two clinical pilots involving three and 12 ambulances, respectively in the Region Västra Götaland in western Sweden [2]. Integrating video analysis of selected NIHSS items into prehospital stroke assessments can serve as a complement to improve the accuracy of the prehospital scales like FAST or mNIHSS. Our previous study showed that video analysis and machine learning (ML) techniques have potential to assist in prehospital assessment of various stroke symptoms, such as facial palsy, paresis of extremities, ataxia and dysarthria (speech disorder) [5]. Incorporating this technology into video consultations [2] could serve as a valuable add-on feature, facilitating more accurate diagnoses and informed triage decisions.

Abnormal eye movement such as gaze palsy (difficulty moving both eyes in a specific direction) are indications of stroke and related neurological impairments [7,8]. One of the NIHSS items evaluates the horizontal eye-movement and if there exists conjugated eye deviation (CED), a condition where both eyes are involuntarily drawn to one side due to brain damage [3,4]. A study reported an accuracy of up to 88% in distinguishing stroke patients from healthy individuals by analyzing dynamic eye features, such as fixation (maintaining steady gaze on a single target) and saccades (rapid and jerky movements that shift the gaze), using eye-tracking and ML techniques [8].

The use of synthetic data is a growing field with a significant potential in providing datasets for the development of ML algorithms in healthcare [9–11]. A study using synthetic data across 19 healthcare datasets found that ML models trained on synthetic data performed nearly as well as those trained on real data, with only a 6–7 percentage point difference in accuracy [9]. In stroke research, platforms like GameSTROKE, which simulate stroke symptoms and CT scans, have demonstrated the potential of synthetic data to enhance prehospital training and assessment [6,12].

This paper introduces a framework for generating synthetic eye movement data to characterize stroke-related eye conditions. Using 3D modeling and video analysis, we create a labeled dataset for training and evaluating ML models. Our approach aims to enhance early stroke assessment in both clinical and prehospital settings.

2. Methods

2.1. Dataset generation

Using Blender [13] and Adobe After Effects [14], 3D models of eyes and heads were created to simulate realistic eye movements and generate synthetic data for ML purposes. A series of video clips (n=73), with eye models integrated into synthetic heads, to reflect eye movement patterns in stroke patients, illustrated in Figure 1. Videos were categorized based on NIHSS horizontal eye movement assessments into three classes: Class 0 (no gaze palsy), Class 1 (partial gaze palsy) and Class 2 (total gaze paralysis). The dataset included 20 videos for Class 0, 26 videos for Class 1 (four videos for each cranial nerve palsy) and 27 videos for Class 2. The videos were reviewed and approved by a stroke neurologist (author K.J.).

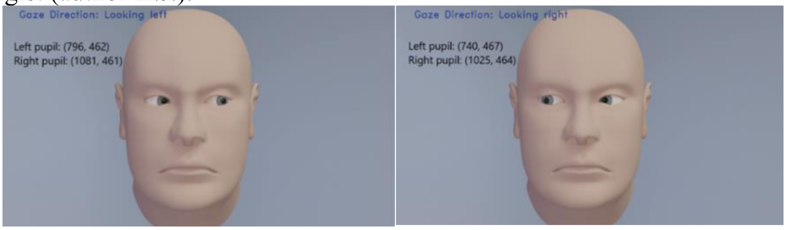

Figure 1. Examples of the generated 3D model, showing direction outputted by GazeTracking.

2.2. Data extraction and processing

GazeTracking software [15] was used to extract relevant eye movement data from the videos. This software uses image processing and ML algorithms to continuously track gaze direction and output X and Y coordinates for each pupil (Figure 1). A program was developed in Python to make the data compatible with ML applications and to process the GazeTracking output. It collects pupil coordinates and gaze direction classification (left, right and center) and then saves them in CSV (Comma-Separated Values) format, with each video frame recorded as a separate row. Then, an additional processing step was performed to calculate the change in coordinates between consecutive frames.

2.3. ML models training and evaluation

Four ML models were selected based on findings from a literature review, representing models suitable for sequence and image tasks with varying complexity: long short-term memory (LSTM), gated recurrent units (GRU), convolutional neural networks (CNN) and support vector machine (SVM) [16]. A 70-30 training-testing split was continuously used, with the test set kept completely separated. For LSTM, GRU and CNN models, the 70% training set was further split into 90–10 for training and validation, allowing early stopping based on validation loss, which was monitored using epoch-loss plots. For SVM model, 3-fold cross-validation was applied within the training set.

The LSTM model included a masking layer to handle irrelevant input, with three bidirectional layers (128, 64 and 64 neurons) and dropout for regularization [16]. The GRU had four layers (256, 128, 64 and 32 units) and outputted a condensed sequence for classification [16]. Both models used categorical cross-entropy as the loss function.

The CNN model included a Gaussian noise layer for robustness, followed by two convolutional layers (64 and 128 filters), with dropout and batch normalization to prevent overfitting [16]. The SVM model used a linear kernel.

The models' performance was finally evaluated on the independent test set using accuracy, precision, sensitivity, specificity, F1-score, false positive rate (FPR) and false negative rate (FNR).

3. Results

The performance of the four models is presented in Table 1. LSTM and GRU showed good, comparable performance across all metrics. SVM performed moderately well, while CNN metrics lagged slightly, especially in accuracy, precision and F1-Score.

Table 1. LSTM, GRU, CNN and SVM models' performance (unit for all metrics is percentage).

Model	Accuracy	Precision	Sensitivity	Specificity	F1-Score	FPR	FNR
LSTM	86.4	84.9	84.9	93.6	84.6	6.4	15.1
GRU	86.4	86.1	86.0	93.5	85.8	6.5	14.0
CNN	72.7	73.0	72.5	85.4	69.5	14.6	27.5
SVM	77.3	77.1	75.7	88.7	76.1	11.3	24.3

Figure 2 presents the confusion matrix for each model. It highlights how LSTM and GRU effectively minimize misclassifications, whereas CNN notably struggles to distinguish partial gaze palsy from other classes.

Figure 2. Confusion matrix of LSTM, GRU, SVM and CNN, evaluated using approximately 22 test samples.

4. Discussion

Early, and fast identification of stroke symptoms is essential for determining appropriate transportation destinations and initiating prenotification to the receiving hospital, enabling faster treatment. In this study, synthetic video data was successfully generated using 3D modeling to mimic eye movements for stroke patients. A neurologist validated the synthetic videos as realistic and clinically relevant, supporting their use in early-stage development without relying on real patient data. LSTM and GRU emerged as top-performing models due to their ability to handle sequential data efficiently, making them well suited for eye movement analysis. SVM offered a simpler, more interpretable approach, though it may lack sensitivity to temporal dependencies. CNN appeared less suited for this task, likely due to its limited handling of temporal patterns.

The main strength of this study is that it provides a proof-of-concept for using synthetic video data in healthcare, especially for stroke-related eye movement. Synthetic data provides a controlled dataset, which could facilitate the development of ML-based healthcare applications while addressing ethical concerns, reducing need for real patient data until later stages in product development where clinical trials may be performed [10]. By minimizing the use of real patient data, this approach facilitates progress in methodology and streamlines the path to clinical application [6]. Real-world validation however remains essential to ensure the model's predictive performance on patient data.

Limitations with this study include the relatively small size and limited variety of the dataset, which consists of synthetic data that may not fully capture the complexity and diversity of real-world stroke cases. Synthetic data is designed to mimic patient data but may not replicate all patient-specific features, potentially impacting model accuracy and generalization to clinical scenarios. The analysis also focused on combined eye movements, and further investigation into right-eye versus left-eye movements could provide valuable insights into detecting abnormalities. The models' generalization to a broader population may also be limited due to the imbalance between synthetic and patient data. Future work could address these limitations and compare patient, synthetic and combined datasets to evaluate model performance across different data sources.

5. Conclusion

This study represents a step towards automated eye movement characterization for stroke patients and demonstrates the potential of using synthetic data for developing healthcare ML models. Synthetic data combined with ML offer a promising approach for early model development. Further research is needed to validate these findings and confirm the effectiveness of this approach in early stroke assessment.

References

[1]. V.L. Feigin, M. Brainin, B. Norrving, S. Martins, R.L. Sacco, et al., World Stroke Organization (WSO): Global Stroke Fact Sheet 2022, *Int J Stroke*. **17** (2022) 18–29. doi:10.1177/17474930211065917.

[2]. S. Candefjord, M. Hagiwara, B.A. Sjöqvist, J.-E. Karlsson, et al., Video support for prehospital stroke consultation: implications for system design and clinical implementation from prehospital simulations, *BMC Med Inform Decis Mak*. **24** (2024) 1–12. doi:10.1186/s12911-024-02539-7.

[3]. NIH Stroke Scale | National Institute of Neurological Disorders and Stroke, (n.d.). https://www.ninds.nih.gov/health-information/stroke/assess-and-treat/nih-stroke-scale.

[4]. T. Brott, H.P. Adams, C.P. Olinger, J.R. Marler, W.G. Barsan, et al., Measurements of acute cerebral infarction: a clinical examination scale., *Stroke*. **20** (1989) 864–870. doi:10.1161/01.STR.20.7.864.

[5]. H. Jalo, A. Borg, E. Thoreström, N. Larsson, M. Lorentzon, O. Tryggvasson, V. Johansson, P. Redfors, et al., Early Characterization of Stroke Using Video Analysis and Machine Learning, in: Emerging Technologies in Healthcare and Medicine, AHFE Open Acces, 2023. doi:10.54941/ahfe1004359.

[6]. E. Christensen, H. Fagerheim Bugge, J. Hagemo, K. Larsen, A.K. Harring, J. Gleditsch, J. Ibsen, M. Guterud, et al., Prehospital stroke diagnostics using three different simulation methods: A pragmatic pilot study, *European Stroke Journal*. **9** (2024) 1016–1024. doi:10.1177/23969873241252564.

[7]. M. Strupp, O. Kremmyda, C. Adamczyk, N. Böttcher, et al., Central ocular motor disorders, including gaze palsy and nystagmus, *Journal of Neurology*. **261** (2014) 542. doi:10.1007/s00415-014-7385-9.

[8]. Q. Lu, J. Deng, Y. Yu, Y. Li, K. Wei, et al.. Yan, Machine learning models for stroke detection by observing the eye-movement features under five-color visual stimuli in traditional Chinese medicine, *Journal of Traditional Chinese Medical Sciences*. **10** (2023) 321–330. doi:10.1016/j.jtcms.2023.06.003.

[9]. D. Rankin, M. Black, R. Bond, J. Wallace, M. Mulvenna, and G. Epelde, Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing, *JMIR Medical Informatics*. **8** (2020) e18910. doi:10.2196/18910.

[10]. J.-F. Rajotte, R. Bergen, D.L. Buckeridge, K. El Emam, R. Ng, et al., Synthetic data as an enabler for machine learning applications in medicine, *iScience*. **25** (2022) 105331. doi:10.1016/j.isci.2022.105331.

[11]. J. Walonoski, M. Kramer, J. Nichols, A. Quina, C. Moesel, D. Hall, et al.. McLachlan, Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record, *J Am Med Inform Assoc*. **25** (2018) 230–238. doi:10.1093/jamia/ocx079.

[12]. A.K.V. Harring, J. Røislien, K. Larsen, M. Guterud, H.F. Bugge, E.C. Sandset, D.V. Kristensen, and M.R. Hov, Gamification of the National Institutes of Health Stroke Scale (NIHSS) for simulation training—a feasibility study, *Advances in Simulation*. **8** (2023) 4. doi:10.1186/s41077-023-00245-4.

[13]. A Review on Different Versions and Interfaces of Blender Software | IEEE Conference Publication | IEEE Xplore, (n.d.). https://ieeexplore.ieee.org/document/10125672.

[14]. Motion graphics software | Adobe After Effects, (n.d.). https://www.adobe.com/products/aftereffects.html.

[15]. A. Lamé, antoinelame/GazeTracking, (2024). https://github.com/antoinelame/GazeTracking.

[16]. A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd ed., O'Reilly Media, Inc., 2019.