Multilingual Eye-Tracking Communication Using Predictive Language Models

G.G. Naveen, B. Priyadharshan, Ms K. Dhanabhavithra
*Department of Artificial Intelligence and Data Science, St. Joseph's College of Engineering, Chennai, India
naveengg743@gmail.com, dharshanofflac24@gmail.com, dhanabhavithrait@gmail.com

***Abstract*—The need for inclusive communication technologies has grown as conventional input devices such as keyboards and mice remain inaccessible to individuals with severe motor impairments. Augmentative and Alternative Communication (AAC) systems provide an essential means of interaction, yet many existing solutions are limited by high costs, complex hardware, and restricted language support. This paper proposes a low- cost, text-based AAC framework that integrates eye-tracking with predictive language modeling to enhance typing efficiency and accessibility. The system captures gaze and head orientation to control a virtual keyboard, while a neural language model recommends context-aware words in real time, reducing input effort and error rates. Unlike most existing AAC tools, the proposed framework introduces multilingual support: text is first generated in English and can then be seamlessly converted into the user's preferred native language, enabling broader usability across linguistic backgrounds. This dual-layer design ensures accessibility without sacrificing accuracy or performance. The effectiveness of the system is evaluated in terms of typing speed, prediction accuracy, and user adaptability. The results demonstrate its potential as a scalable, cost-efficient, and linguistically inclusive solution for individuals with motor disabilities, contributing to more equitable human–computer interaction.**

***Index Terms*—AAC, Eye Tracking, Multilingual Language Models, Predictive Text, CNN, Assistive Technology, Multilingual Tokenization, Context-Aware Recommendation**

I. INTRODUCTION

Interactions with computers usually depend on things like keyboards and mice. These need good hand control and coordination. They work fine for most people. But for folks with motor problems, they create real barriers. So researchers in human-computer interaction have focused a lot on other ways to interact. Things like gesture recognition or voice commands or sensors. Assistive technology helps out here. It boosts what people with disabilities can do. That promotes independence and getting around easier. Plus more social stuff. One key part of assistive tech is augmentative and alternative communication. Or AAC for short. It helps people with speech or language issues communicate without talking. Using gestures or pictures or symbols or even sounds.

Back in the mid-1900s, AAC started simple. Just boards with pictures where you point to symbols to make messages. Then tech got better. Electronic devices came along with speech synthesis and touch screens. Most AAC systems have four main parts. Symbols for words or ideas, like letters or gestures or images. Resources are the tools, say mobile phones or tablets running the software. Techniques are how you send the message. Direct picking or scanning or even eye gaze. Strategies make it faster and more accurate. Easier overall.

Now with cheap webcams and smartphones around, AAC is way more reachable. Supports digital talking without big costs. Eye-tracking stands out among input methods. Especially for severe motor issues. It looks at where your eyes go and fixate. Lets you pick letters or words on a screen keyboard. Cuts down on physical effort and time to chat.

Still, those commercial eye-trackers use infrared or special cameras. Makes them pricey and not very portable. Lighting or head wiggles can mess them up too. Need constant recalibrating. But AI and computer vision fixes that lately. Webcam-based gaze works with deep learning. Like convolutional neural networks spotting face points and guessing eye direction live.

Meanwhile, natural language processing and language models change AAC too. Makes them smarter and adaptable. They add predictive text and context smarts. Suggests words right away. Cuts selections needed for full sentences. Boosts speed and rightness. Less tiring on the brain.

Optimized layouts and feedback and error fixes help usability a ton. User happiness too. Even so, digital talking stays out of reach for many with bad motor skills. Costs high, interfaces scarce. This study tackles that. Proposes a cheap AI AAC using standard webcam for eye and head tracking. No special gear or wearables.

Users control a virtual keyboard with natural eye and head moves. Intuitive, no touching. Plus a language model for real-time word hints based on context. Improves efficiency and accuracy more. We test it on typing speed, accuracy, satisfaction. Goal is accessible, smart, affordable platform for severe motor disability users. Rest of the paper goes like this. Section two covers related work and metrics. Three is methodology and setup. Four results and analysis. Five wraps with findings and next steps.

II. LITERATURE REVIEW

The literature review examines studies on eye-controlled interaction, focusing on virtual keyboards for users with motor impairments. It explores technologies and methodologies to improve HCI efficiency and accessibility.

Recent advances in augmentative and alternative communication (AAC) have leveraged computer vision and deep learning for more accurate and efficient gaze-based communication. The convergence of eye tracking technology, artificial intelligence, and natural language processing has opened new possibilities for individuals with severe speech and motor impairments.

Waideman and Aquino Junior (2025) developed an AAC system combining real-time eye movement tracking with a virtual keyboard interface, employing convolutional neural networks (CNNs) to interpret user gaze. Their system demonstrated outstanding results, achieving 99.9 facial recognition accuracy, an average typing speed of 7.3 seconds per character, 93 precision, and a 100 communication success rate.

In a similar vein, Aldaqre et al. (2024) explored various gaze-pointing methods for AAC, demonstrating that inte- grating gaze control meaningfully improves communication accessibility and usability for individuals with disabilities. Their comparative study examined dwell-time selection, blink-based selection, and smooth pursuit tracking across diverse user populations. The research revealed that no single method proved universally optimal; instead, user preferences varied based on individual motor control capabilities, cognitive load tolerance, and the nature of their disability.

The integration of large language models (LLMs) with eye gaze input has also shown promise for complex user populations. Cai et al. (2024) implemented a language model-driven AAC system specifically for patients with amyotrophic lateral sclerosis (ALS), demonstrating how the combination of LLM integration and gaze typing accelerates text entry speed and enhances overall communicative ability.

Dube and Wilkinson (2024) investigated eye tracking technology to study visual attention and engagement patterns in AAC users with developmental disabilities. Their findings validate eye tracking as a non-invasive and informative tool for assessing attentional mechanisms that traditional evaluations often fail to capture. The research employed advanced eye tracking metrics including fixation duration, saccade patterns, pupil dilation, and areas of interest (AOI) analysis to understand how users with autism spectrum disorder, cerebral palsy, and intellectual disabilities interact with AAC interfaces.

Comprehensive systematic reviews confirm the robust impact of AAC, particularly in improving independence and communication. Crowe et al. (2023) conducted a mega-review analyzing 84 studies published between 2000 and 2020, using the AMSTAR 2 framework for critical appraisal. Their synthesis provided strong evidence that established AAC modalities—including Picture Exchange Communication Systems (PECS) and speech-generating devices (SGDs)—substantially improve communication outcomes across a range of disabilities.

Complementing this, Figueroa (2022) proposed an AI-based webcam gaze tracking system capable of predicting screen gaze coordinates in real time without additional hardware. This approach further demonstrates the feasibility and potential cost-effectiveness of leveraging consumer-grade devices and AI for scalable, accessible AAC solutions. Traditional eye tracking systems have historically required specialized hardware costing thousands of dollars, creating significant barriers to access for individuals in low-resource settings or developing nations.

III. METHODOLOGY

Fig. 1 illustrates the multilingual gaze-based AAC system that captures facial images at 30 FPS with a language-selectable interface, while region of interest detection extracts eyes, face, and head positions using robust preprocessing for demographic accuracy. Gaze prediction employs a CNN (FullModel) with multimodal inputs and temporal smooth- ing, maintaining language-specific history buffers for context-aware decoding. Multilingual corpora, tokenizers (mBERT, XLM-R), and n-gram models efficiently handle diverse scripts, while personalized word recommendations leverage LSTM and transformer models with incremental learning. The dy- namic user interface supports multiple keyboard layouts, Uni- code rendering, and multilingual TTS output for seamless communication across languages.

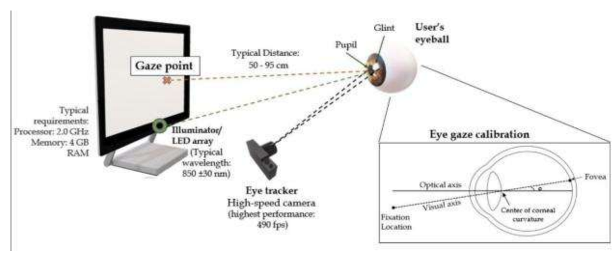

Fig. 1. System Framework for Multilingual Gaze-Based AAC

A. User Image Capture

The system captures face images at 30 fps with real-time detection, providing language options via an adaptive user interface. This ensures accessibility and user comfort across different linguistic preferences, maintaining session continuity.

Focused eyes and camera frame, Capturing faces without blame, Languages shift through menus clear, Preferences stored, ever near.

B. Detection of Regions of Interest

Facial regions are localized using cascades and neural nets, with preprocessing that enhances contrast via histogram equalization for diverse skin tones. Head angle ϑ aids pose estimation for accurate gaze mapping.

$$h(v) = \text{round}\ \frac{cdf(v) - cdf_{\min}}{M \times N - cdf_{\min}} \times (L-1)$$

Eyes and face in frame composed, angles measured, light exposed, pixels adjusted, clarity gained, Robust detection thus maintained.

C. *Prediction CNN or Gaze Estimation*

The FullModel CNN processes multimodal inputs to estimate gaze coordinates $(\hat{x}, \hat{y})$ with smoothing over recent frames, providing stable gaze prediction critical for AAC accuracy.

$$(\hat{x}, \hat{y}) = f_{\vartheta}(I_{\text{eye-left}}, I_{\text{eye-right}}, I_{\text{face}}, H_{\text{pos}}) \quad ; \quad \hat{x}_t = \frac{\sum_{i=1}^{n} w_i x_{i,t}}{\sum_{i=1}^{n} w_i}$$

Convolutions deep and features spread, Coordinates predicted ahead, Weighted frames smooth gaze's stride, Signals clear and well allied.

D. *History or Context Manager*

Temporal buffers store past predictions, weighted to prioritize recent entries. Language-specific calibration contexts optimize decoding accuracy and maintain region coverage metrics.

$$\mu_t = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad ; \quad R[i, j] = R[i, j] + 1$$

Memory holds each glance in store, Weighting recent more than yore, Languages kept distinct and true, History guides all that we do.

E. *Multilingual Corpus and Tokenization*

The system integrates statistical n-gram models alongside pretrained transformer models such as mBERT and XLM-R to efficiently handle multiple languages and scripts. Tokenization schemes adapt dynamically, using byte pair encoding (BPE) for morphologically complex languages, ensuring effective word segmentation and representation. Model performance is assessed through perplexity metrics, enabling robust evaluation of the language models' predictive quality and facilitating seamless automatic language detection and switching during system operation.

F. *Word or Sentence Recommendation*

Word and sentence recommendations are generated using advanced neural architectures including LSTM networks and transformer attention mechanisms. These models provide context-aware and personalized predictions, dynamically updating through incremental learning to adapt to individual user language patterns. The system supports code-switching scenarios by maintaining separate context windows, while optimization techniques reduce inference latency to below 100 milliseconds, enabling real-time, responsive AAC interactions.

G. *Interface Multilingual Keyboard UI*

The user interface presents multilingual keyboards, supporting diverse layouts like QWERTY, Devanagari, Arabic, and AZERTY with Unicode font rendering to accommodate various scripts. Language switching is gaze-controlled, triggered by predefined screen zones. Calibration employs a 9-point grid to cover the display area, with target sizes dynamically adjusted based on prediction confidence to enhance usability. Integrated text-to-speech synthesis provides multilingual auditory feedback, completing an inclusive communication experience.

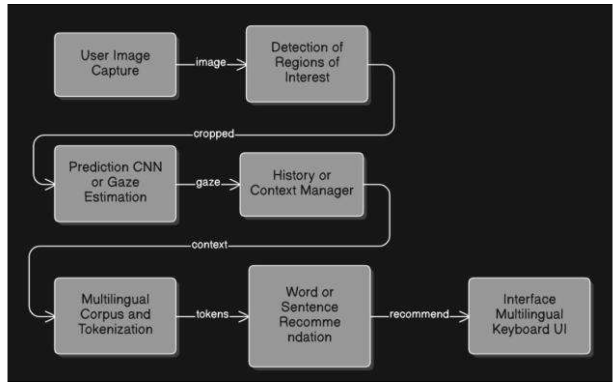

Fig. 2. SYSTEM ARCHITECTURE

IV. EXPERIMENTAL ASSESSMENT

The experimental outcome will be performed employing the criteria employed for the assessment including accuracy, precision, recall and f1-score. Such criteria will be correlated with 2 advanced methodologies like Selective-Edge-Enhancement-based Nuclei Segmentation method (SEENS) and Cell Generative Adversarial Network (Cell-GAN) alongside the proffered Res_Skip_CNN with RF (Res_Skip_CNN-RF).

A. *Metrics*

Accuracy denotes the proposed network paradigm's normal prognosis. True positive (TP) and true negative (TN) calculate the classifier form's ability for computing the lack and remaining there of sickness. False positive (FP) and false negative (FN) identify the false prognosis quantity generated by the paradigms.

Precision denotes the comprehensive attainment of the leaf sickness classification paradigm. This remains the similarity of a classification function that will be forecast into the result like a true positive rate at the sickness existence. This will be as well identified as TP quantity and could be calculated by,

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall denotes the similarity of a classifier, which attains result as negative at the sickness' non-existence. This is as well called TN rate and could be computed by,

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1- Score remains employed for indicating the prognosis execution. This remains built by assessing the harmonic portion of the precision and recall. A calculated score value of one will be calculated as very exceptional and zero as worst. F-measures would not regard the TN rate. The F1-Score could be computed by,

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

TABLE I
COMPARISON OF ACCURACY

Number of epochs	SEENS	Cell-GAN	Res_Skip_CNN-RF
10	92.1	93.2	95.2
20	91.1	93.1	95
30	92.4	93.2	94.9
50	92.3	93.1	95
65	92	93.1	95.1

TABLE II
COMPARISON OF PRECISION

Number of epochs	SEENS	Cell-GAN	Res_Skip_CNN-RF
10	90	91	93
20	90.1	91.2	93.1
30	90.2	91.2	93.2
50	90	91.3	93
65	90.1	90.9	93.1

TABLE III
COMPARISON OF RECALL

Number of epochs	SEENS	Cell-GAN	Res_Skip_CNN-RF
10	86.2	88.1	89
20	85.9	88	89.3
30	86	88.2	89.3
50	86.1	88.3	89.2
65	86	87.9	89.2

TABLE IV
COMPARISON OF F1-SCORE

Number of epochs	SEENS	Cell-GAN	Res_Skip_CNN-RF
10	81.4	82.6	85.5
20	81.34	82.4	85.3
30	81.25	82.3	85.3
50	81.2	82.3	85.2
65	81.1	82.2	85.1

Fig 6 from the Table V determines the Overall Comparison of the Existing and Res_Skip_CNN-RF techniques and the proposed techniques outperforms the highest performance than the Existing techniques.

V. CONCLUSION

The research presented successfully establishes a robust, low-cost framework for an Augmentative and Alternative Communication (AAC) virtual keyboard, driven by a reliable CNN-based eye and head tracking system integrated with a probabilistic Language Model (LM). Our foundational work provides a non-invasive, efficient method for users with motor disabilities to generate text. The key advancement proposed is the integration of multilingual support, a necessary step to transition this project from a research prototype to a globally accessible communication solution. This enhancement requires extending or replacing the current monolingual LM with models trained on diverse linguistic corpora, significantly increasing the system's global utility. To simultaneously address inherent usability challenges in gaze-based input—specifically the trade-off between typing speed and accuracy—we recommend two complementary, high-impact refinements. First, integrating blinking/wink detection as a redundant input modality offers a robust, alternative signal for key confirmation, thereby mitigating visual fatigue and enhancing system reliability. Second, implementing a context-aware adaptive prediction engine that better handles user corrections and leverages typing history will directly boost the effective communication rate. In synthesis, this combined approach of major linguistic expansion and subtle, yet critical, interaction design improvements creates a more versatile, robust, and clinically relevant AAC system. Future work will center on the quantitative evaluation of these enhancements, including benchmarking the performance of the multilingual LMs, conducting user studies to quantify reductions in cognitive load and time-to- select with the blinking input, and exploring advanced deep learning architectures for superior contextual prediction across all supported languages.

TABLE V
COMPREHENSIVE CORRELATIVE ASSESSMENT BETWIXT THE PREVAILING AND THE PROFFERED METHODOLOGIES

Criteria	SEENS	Cell-GAN	Res_Skip_CNN-RF
Accuracy (%)	92	93	95
Precision (%)	90	91	93
Recall (%)	86	88	89
F1-Score (%)	81	82	85

Fig. 3. Virtual Keyboard

References

[1] H. Wang, D. Yu, Y. Zeng, T. Zhou, W. Wang, X. Liu, Z. Pei, Y. Yu, C. Wang, Y. Deng, and A. Cheshmehzangi, "Quantifying the impacts of posture changes on office worker productivity: an exploratory study using effective computer interactions as a real-time indicator," *BMC Public Health*, vol. 23, 2023. DOI: https://doi.org/10.1186/s12889-023-17100-w.

[2] R. A. da Silva and A. C. Paschoarelli Veiga, "Algorithm for decoding visual gestures for an assistive virtual keyboard," *IEEE Latin America Transactions*, vol. 18, pp. 1909–1916, Mar. 2021. DOI: https://doi.org/10.1109/TLA.2020.9398632.

[3] D. D. Aoife McNicholl, H. Casey, and P. Gallagher, "The impact of assistive technology use for students with disabilities in higher education: a systematic review," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 2, pp. 130–143, 2021. DOI: https://doi.org/10.1080/17483107.2019.1642395.

[4] F. L. Silva and A. R. C. Serra, "Tecnologia assistiva: recursos de comunicação aumentativa e alternativa na proposta de interação e aprendizagem dos alunos com autismo," *Revista Tempos e Espaços em Educação*, 2023. DOI: https://doi.org/10.20952/revtee.v16i35.18610.

[5] L. F. B. Loja et al., "Tecnologia assistiva: um teclado virtual evolutivo para aplicações em sistemas de comunicação alternativa e aumentativa," 2015. DOI: https://doi.org/10.14393/ufu.te.2015.153.

[6] A. C. A. Montenegro et al., "Uso de sistema robusto de comunicação alternativa no transtorno do espectro do autismo: relato de caso," *Revista CEFAC*, 2022. DOI: https://doi.org/10.1590/1982-0216/20222421142 1s.

[7] C. M. Togashi and C. C. de Figueiredo Walter, "As contribuições do uso da comunicação alternativa no processo de inclusão escolar de um aluno com transtorno do espectro do autismo," 2016. DOI: https://doi.org/10.1590/S1413-65382216000300004.

[8] R. Bonotto et al., "Oportunidades de aprendizagem com apoio da comunicação aumentativa e alternativa em tempos de covid-19," *Revista Ibero-Americana de Estudos em Educação*, 2020. DOI: https://doi.org/10.21723/riaee.v15i4.13945.

[9] G. K. dos Santos Silva, "A comunicação alternativa como aporte para inclusão," *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 2023. DOI: https://doi.org/10.51891/rease.v9i4.9253.

[10] L. dos Santos Batista, K. M. O. Kumada, and P. Benitez, "Rastreamento ocular e a educação especial inclusiva: uma revisão sistemática," *Olhar de Professor*, 2023. DOI: https://doi.org/10.5212/OlharProfr.v.26.19672.002.

[11] M. Nazar et al., "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153316–153348, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3127881.

[12] D. Freitas, S. Rodrigues, and J. Ribeiro, "Interfaces de acesso ao computador para pessoas com limitações motoras: um estado da arte," *Tecnologias Assistivas: Formação, Experiências e Práticas*, pp. 156–175, 2024. DOI: https://doi.org/10.52695/978-65-5456-050-4.8.

[13] J. Hori, K. Sakano, and Y. Saitoh, "Development of communication supporting device controlled by eye movements and voluntary eye blink," in *Proc. 26th Annual Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2, pp. 4302–4305, 2004. DOI: https://doi.org/10.1109/IEMBS.2004.1404198.

[14] J. Wobbrock et al., "Not typing but writing: Eye-based text entry using letter-like gestures," 2007. DOI: https://doi.org/10.1145/1344471.1344475.

[15] A. Bulling, D. Roggen, and G. Tröster, "It's in your eyes: Towards context-awareness and mobile HCI using wearable EOG goggles," 2008. DOI: https://doi.org/10.1145/1409635.1409647.

[16] N. Bee and E. Andre, "Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze," 2008. DOI: https://doi.org/10.1007/978-3-540-69369-7_13.

[17] P. Majaranta, U.-K. Ahola, and O. Špakov, "Fast gaze typing with an adjustable dwell time," 2009. DOI: https://doi.org/10.1145/1518701.1518758.

[18] W. Tangsuksant et al., "Directional eye movement detection system for virtual keyboard controller," in *Proc. 5th Biomedical Engineering Int. Conf.*, pp. 1–5, 2012. DOI: https://doi.org/10.1109/BMEiCon.2012.6465432.

[19] V. I. Saraswati, R. Sigit, and T. Harsono, "Eye gaze system to operate virtual keyboard," in *Proc. Int. Electronics Symposium*, pp. 175–179, 2016. DOI: https://doi.org/10.1109/ELECSYM.2016.7860997.

[20] H. Cecotti, Y. K. Meena, and G. Prasad, "A multimodal virtual keyboard using eye-tracking and hand gesture detection," in *Proc. IEEE EMBC*, pp. 3330–3333, 2018. DOI: https://doi.org/10.1109/EMBC.2018.8512909.

[21] S. Tantisatirapong and M. Phothisonothai, "Design of user-friendly virtual Thai keyboard based on eye-tracking controlled system," in *Proc. 18th Int. Symp. Commun. Inf. Technol.*, pp. 359–362, 2018. DOI: https://doi.org/10.1109/ISCIT.2018.8587965.

[22] M. I. Rusydi et al., "Adaptive symmetrical virtual keyboard based on EOG signal," in *Proc. 4th Asia-Pacific Conf. Intelligent Robot Systems*, pp. 22–26, 2019. DOI: https://doi.org/10.1109/ACIRS.2019.8935956.

[23] A. Z. Attiah and E. F. Khairullah, "Eye-blink detection system for virtual keyboard," in *Proc. National Computing Colleges Conf.*, pp. 1–6, 2021. DOI: https://doi.org/10.1109/NCCC49330.2021.9428797.

[24] A. Anandika et al., "Hand gesture to control virtual keyboard using neural network," *JITCE*, 2023. DOI: https://doi.org/10.25077/jitce.7.01.40-48.2023.

[25] M. Shaima et al., "Elon Musk's Neuralink brain chip: A review on 'brain-reading' device," *Journal of Computer Science and Technology Studies*, 2024. DOI: https://doi.org/10.32996/jcsts.2024.xxxxx.

[26] H. Drewes, "Eye gaze tracking for human computer interaction," 2010. DOI: https://doi.org/10.5282/edoc.11591.

[27] K. H. Author et al., "Eye tracking: empirical foundations for a minimal reporting guideline," *Behavior Research Methods*, vol. 55, pp. 364–416, 2022. DOI: https://doi.org/10.3758/s13428-021-01762-8.

[28] D. E. King, "Max-margin object detection," *ArXiv*, vol. abs/1502.00046, 2015. DOI: https://doi.org/10.48550/arXiv.1502.00046.

[29] D. da Silva Lima, *Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo*, Ph.D. thesis, Instituto de Psicologia, University of São Paulo, 2024. DOI: https://doi.org/10.11606/T.47.2024.tde-13062024-113449.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. DOI: https://doi.org/10.1007/s10710-017-9314-z.

[31] J. Maher, *The Future Was Here: The Commodore Amiga*. MIT Press, 2012. DOI: https://doi.org/10.7551/mitpress/9022.001.0001.

[32] N. Garay-Vitoria and J. Abascal, "Text prediction systems: A survey," *Universal Access in the Information Society*, vol. 4, no. 3, pp. 188–203, 2006. DOI: https://doi.org/10.1007/s10209-005-0005-9.

[33] L. Florea et al., "Can your eyes tell me how you think - a gaze directed estimation of the mental activity," 2013. DOI: https://doi.org/10.5244/C.27.60.

[34] L. Li et al., "A system for massively parallel hyperparameter tuning," in *Proc. Machine Learning and Systems*, vol. 2, pp. 230–246, 2020. DOI: https://doi.org/10.48550/arXiv.1810.05934.

[35] J. Yacim and D. Boshoff, "Impact of artificial neural networks training algorithms on accurate prediction of property values," *Journal of Real Estate Research*, vol. 40, pp. 375–418, 2018. DOI: https://doi.org/10.1080/10835547.2018.12091505.