Received: 3 December 2023 | Revised: 1 October 2024 | Accepted: 4 October 2024

DOI: 10.1111/epi.18155

Epilepsia®

RESEARCH ARTICLE

Evaluating novel in silico tools for accurate pathogenicity classification in epilepsy-associated genetic missense variants

Ludovica Montanucci[1,2] | **Tobias Brünger**[3] | **Christian M. Boßelmann**[4,5] | **Alina Ivaniuk**[4] | **Eduardo Pérez-Palma**[6] | **Samden Lhatoo**[1] | **Costin Leu**[1,2,4] | **Dennis Lal**[1,2,3,4,5,7]

[1]Department of Neurology, McGovern Medical School at UTHealth, Houston, Texas, USA

[2]Center for Neurogenetics, UTHealth Houston, Houston, Texas, USA

[3]Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany

[4]Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA

[5]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

[6]Universidad del Desarrollo, Genética y Genómica, Facultad de Medicina Clínica Alemana, Santiago, Chile

[7]Stanley Center of Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

Correspondence
Dennis Lal, Department of Neurology, McGovern Medical School, UTHealth Houston, 1133, John Freeman Blvd, Houston, TX 77030, USA.
Email: dennis.lal@uth.tmc.edu

Abstract

Objective: Determining the pathogenicity of missense variants in clinical genetic tests for individuals with epilepsy is crucial for guiding personalized treatment. However, achieving a definitive pathogenic classification remains challenging, with most missense variants still classified as variants of uncertain significance (VUS) and with the availability of many computational tools which may provide conflicting predictions. Here, we aim to evaluate the performance of state-of-the-art computational tools in pathogenicity prediction of missense variants in epilepsy-associated genes. This will assist in selecting the most appropriate tool and critically assess their use in clinical setting.

Methods: We assessed the performance of nine in silico pathogenicity prediction tools for missense variants in epilepsy-associated genes on three carefully curated data sets. The first two data sets comprise missense variants in epilepsy associated genes that have been uploaded to ClinVar in the last year and were, therefore, not part of the training set of any of the nine considered tools. These two data sets are based on two different lists of epilepsy-associated genes and comprise ~700 and ~250 missense variants, respectively. The third data set includes ~400 missense variants within epilepsy-associated genes for which the functional effects have been determined experimentally and are therefore used here to infer pathogenicity. These three data sets represent the best available approximation to blind and independent test sets.

Results: Among the nine assessed tools, AlphaMissense (area under the curve [AUC]: .93, .88, and .95) and REVEL (AUC: .93, .88, and .93) showed the best classification performance, also outperforming other tools in the number of classified variants.

Significance: We show which recently developed prediction tools achieve higher performance in epilepsy-associated genes and should be integrated, therefore, into the American College of Medical Genetics and Genomics/Association

of Molecular Pathology (AGMC/AMP) variant classification process. Periodic reevaluation of genetic test results with newly developed or updated tools should be incorporated into standard clinical practice to improve diagnostic yield and better inform precision medicine.

KEYWORDS

AlphaMissense, ClinVar, epilepsy, epilepsy-associated genes, experimentally tested variants, variant pathogenicity classification

1 | INTRODUCTION

Pathogenicity classification of genetic variants in epilepsy is extremely important for proper diagnosis of genetic epilepsies, to guide treatment, as well as to enable the discovery of possible pharmacological and therapeutic targets. However, variants of uncertain significance (VUS) are reported in 22.5%–32.6% of clinical genetic diagnostics.[1] Previous assessments of variant classification tools showed limited accuracy on clinical data sets[2] and specifically in epilepsy, more than half of the variants with discordant interpretations were located in genes that have therapeutic implications.[3] Over the last two decades, bioinformatic genetic variant classification tools have undergone a profound transformation, evolving from basic algorithms to highly sophisticated platforms that integrate diverse data sources and cutting-edge machine learning techniques. In the early 2000s, tools relied primarily on simple sequence alignment and conservation metrics to predict variant pathogenicity. Early tools like PolyPhen-2 and SIFT laid the groundwork by predicting the impact of missense mutations based on sequence homology and physical properties of amino acids. The last decade has marked a dramatic evolution with the integration of large-scale genomic databases, improved computational models, and advanced interpretative frameworks. Recent advancements in the aggregation of large-scale clinical and population-level variant data, such as ClinVar and the Genome Aggregation Database (gnomAD), have provided a much richer context for variant interpretation. The adoption of machine learning and artificial intelligence (AI), as exemplified by tools like CADD[4] and REVEL,[5] has revolutionized the field, improving predictive accuracy by adopting more complex models and incorporating multifaceted genomic annotations and functional data. Most of these recent tools are supervised, meaning they were developed using variants with known clinical labels (ranging from benign to pathogenic), such as those present in ClinVar. This makes it difficult to find independent and unbiased test sets for the evaluation of their performance. Unsupervised prediction tools, which do not rely on clinical labels in their development, usually achieve lower performance, with the exception of EVE.[6] In the last 2 years, tools like PrimateAI 2.0 and AlphaMissense[7] have emerged, leveraging deep learning and extensive evolutionary data to predict variant pathogenicity with unprecedented accuracy and, for AlphaMissense, drastically reduce the number of VUS. These latest tools incorporate comprehensive data sets from diverse species and integrate functional and structural genomics data, including the AI-based AlphaFold, pushing the boundaries of precision in genetic variant classification even further.

Key points

- Correct interpretation of clinical genetic tests is crucial for diagnosing genetic epilepsies and guiding personalized treatments.
- The ability of state-of-the-art computational tools to provide accurate clinical pathogenicity predictions in epilepsy-associated genes has not yet been assessed.
- We derived three data sets of missense variants in epilepsy-associated genes and assessed the performance of major state-of-the-art computational tools.

The continuous development and updating of guidelines, such as the ACMG/AMP standards, have standardized variant interpretation practices, ensuring greater consistency and reliability in clinical settings. An additional layer of complexity involves how to integrate these prediction tools within the ACMG guidelines. In the 2015 ACMG guidelines it was established that prediction tools could provide only "supporting" evidence. However, recently, a calibration system[8] has been proposed that quantifies the strength of evidence for major computational tools, with some predictors also reaching strong evidence when certain thresholds of individual scores are met.

Collectively, these advancements have dramatically enhanced our ability to understand and manage genetic diseases, ushering in a new era of personalized medicine.

Here we focus on variants that are likely to be relevant for genetic epilepsies and derive three well-selected epilepsy variant data sets and test the classification performances and the number of classified variants of the major state-of-the-art pathogenicity prediction tools to assess their utility in diagnosing genetic epilepsies.

2 | METHODS

2.1 | Selection of epilepsy-associated genes

Because a definitive and universally accepted list of epilepsy-associated genes is not available, we defined two different lists of epilepsy-associated genes derived according to two different criteria. The first list of epilepsy-associated genes was manually curated by Macnee et al. (2023)[9] and comprises 143 genes (Macnee epilepsy genes). The second list of epilepsy-associated genes (ClinGen-based epilepsy genes) comprises the 100 genes for which an established Clincal Genome resource (ClinGen) gene-disease validity[10] association was present with one or more of 635 epilepsy Mondo Disease Ontology (Mondo) terms[11] derived by mapping Mondo ontology to the Seizure Classification of the International League Against Epilepsy (ILAE).[12] Notably, the overlap of the two gene lists is only ~50 genes, as genes contributing to a wider spectrum of neurological diseases are present in the list of Macnee et al. but not in the ClinGen-based list, where only genes strictly associated with epilepsy were reported.

2.2 | ClinVar-derived variant data sets

To obtain variants in epilepsy-associated genes, we downloaded two releases of the ClinVar database (XML version): release 06-10-2022 (ClinVar2022) and release 01-10-2023 (ClinVar2023). We selected all the variants located in the epilepsy-associated genes that were present in ClinVar2023 but not in ClinVar2022, that is, we selected the variants in epilepsy-associated genes that were added to the database after the development of AlphaMissense and all previously developed prediction tools. Two additional filters were applied: (1) we retained only genes for which at least one benign and one pathogenic variant were reported, to account for the potential bias due to the fact that most genes have either predominantly pathogenic or predominantly benign variants; (2) we retained only variants with a ClinVar review status ≥1 star; and (3) we removed variants for which another variant at the same sequence position of the same gene was present in ClinVar2022, to account for the fact that variants at the same position have highly correlated pathogenicity. This second filter constitutes a very stringent requirement, which is usually not adopted by most computational tools, except, to the best of our knowledge, by AlphaMissense. This procedure yielded two different variant data sets, one for each epilepsy gene list. For the Macnee et al. gene list, we derived a data set consisting of 708 variants from 73 genes, of which 60 were pathogenic, 180 likely pathogenic, 106 benign, and 362 were likely benign, in ClinVar. Of the 708 variants of the Macnee data set, 702 have a clinical review status of 1 star ("criteria provided, single submitter,") two have a status of "criteria provided, multiple submitters, no conflicts," and four have a status of "reviewed by expert panel." For the ClinGen-based gene list, we derived a data set consisting of 255 variants from 34 genes, of which 34 classified as pathogenic, 79 as likely pathogenic, 48 as benign, and 94 as likely benign, in ClinVar. Of the 255 variants of the ClinGen-derived data set, 251 have a clinical review status of 1 star ("criteria provided, single submitter"), and four have a status of "reviewed by expert panel." A flowchart of the data set collection is shown in Figure 1.

2.3 | Experimentally-derived variant data set

This data set consisted of variants from three different families of ion channels (sodium channels, potassium

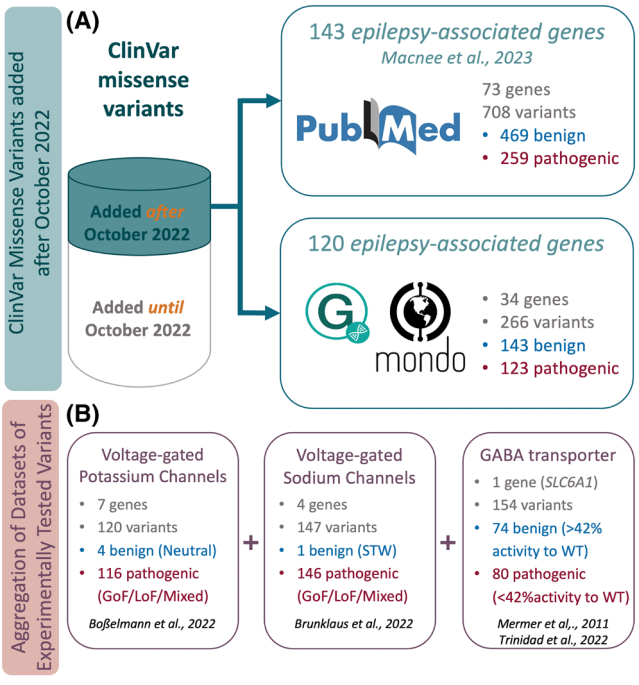

FIGURE 1 Flowchart of the aggregation of variant data from ClinVar (A) and from the experimental data set (B).

channels, and γ-aminobutyric acid [GABA] transporters). For each variant, the functional effect on ion channel activity was measured experimentally. Among these available genes, we selected only those present in either the Macnee et al. or the ClinGen-based epilepsy gene list. A flowchart of the data set collection of the experimental data set is shown in Figure 1. These data sets include variants identified through literature reviews of more than 100 scientific papers,[13–15] which were either potentially pathogenic or confirmed as disease-causing, and, for this reason, these variants subsequently underwent experimental functional testing. Although these variants did not necessarily undergo ACMG classification, most have been derived from patients and, therefore, these data sets show an over-representation of pathogenic variants relative to benign ones. Although the variants in these data sets are potentially pathogenic due to their patient origins, in this work, a variant was considered pathogenic if in vitro assays showed altered function of the mutant protein. Conversely, a variant was considered benign if the in vitro assay showed no functional difference compared to wild-type protein.

2.3.1 | Potassium channel dataset

This data set comprises 120 missense variants from seven genes. (*KCNA2*, *KCNB1*, *KCNC1*, *KCNC2*, *KCNH1*, *KCNQ2*, and *KCNQ3*) encoding human potassium channels involved in epilepsy. The experimental functional classification of these variants was curated by Boßelmann et al. (2022).[14] Of these 120 variants, which were functionally tested by in vitro electrophysiology, 4 were found to have no effect on protein activity and were consequently considered benign; 116 variants were found to have either a loss of function (LoF) or gain of function (GoF) effect or a mixed effect and were thus considered pathogenic. The remaining 11 variants with an unclear effect were removed from subsequent analysis.

2.3.2 | Sodium channel data set

This data set comprises 147 missense variants from four genes (*SCN1A*, *SCN2A*, *SCN3A*, and *SCN8A*) encoding voltage-gated sodium channel proteins involved in epilepsy. The experimental functional classification of these variants was curated by Brunklaus et al. (2022).[15] Of these 147 variants, 146 were found to have a GoF, LoF, LoF/mixed, or mixed effect. These effects were considered pathogenic, whereas the one variant whose experimental readout was found to be "similar to wild type" was considered benign.

2.3.3 | GABA transporter data set

This data set comprises 154 missense variants from the *SLC6A1* gene encoding the GABA transporter type 1. The functional impact of the 154 variants on the protein activity was measured experimentally and reported as percentage of wild-type activity.[13,16] Variants whose residual activity was <42.8% with respect to wild-type activity were found to be LoF,[13,16] and therefore are here considered pathogenic, whereas the remaining variants were considered as benign. These correspond to 80 pathogenic variants and 74 benign variants.

2.4 | Computation of pathogenicity scores and classification

For each variant, the following scores were annotated using ANNOVAR[17] with custom databases for deleteriousness assessment: REVEL,[5] EVE,[6] MutationAssessor,[18] CADD,[4] PrimateAI,[19] MPC, M-CAP,[20] and PolyPhen2.[21] AlphaMissense predictions were derived from its repository on Zenodo (https://zenodo.org/records/8208688), and only the main predictions relative to the MANE transcripts were considered.

To compute the number of variants predicted as VUS by each prediction tool, it is necessary to determine the classification into the three classes of pathogenic, benign, or ambiguous in case of low prediction confidence. This was not available for all the considered prediction tools. For REVEL, PrimateAI, and MPC, a variant was considered pathogenic or benign if its score provided *supporting evidence* for that class, according to the thresholds established by the calibrations of the PP3/BP4 ACMG criteria,[8] whereas it was considered ambiguous if its score fell between these classifications. For AlphaMissense, it was provided along with the score. Since EVE has not undergone calibration to establish score cutoffs aligned with evidence strength as per ACMG guidelines,[8] we adopted the thresholds provided by the developers of the EVE method and reported in Frazer et al. (2021),[6] which involved excluding the top 25% of the most uncertain predictions. Therefore, we assessed the top 75% confident variants as pathogenic or benign and the remaining ones as ambiguous. MutationAssessor and M-CAP were excluded from this analysis because they were not included in the calibration of PP3/BP4 ACMG criteria,[8] and the developers did

not provide thresholds or criteria for differentiating ambiguous predictions.

3 | RESULTS

3.1 | Pathogenicity prediction performances on variants from ClinVar-derived data sets in epilepsy-associated genes

The variant selection process generated two data sets of variants derived from ClinVar. The performance of the nine considered prediction tools on these two ClinVar-derived data sets are shown in Figure 2A (for the variant data set based on Macnee et al.) and 2b (for the variant dataset based on ClinGen) through the receiver-operating characteristic (ROC) curves (the higher the curve, the better the prediction) and the area under the curve (AUC) value. AUC ranges from 0, when all predictions are wrong, to 1 when all predictions are correct. AlphaMissense and REVEL achieved the best classification performance, both reaching AUC .93 on the Macnee et al. data set and AUC 0.88 on the ClinGen data set. MutationAssessor (AUC .90 and .84) and EVE (AUC 0.84 and 0.87) were the second-best computational tools. It is important to note that most prediction tools do not provide predictions for all possible variants, due to variants being annotated on non-canonical isoforms, or due to characteristics inherent to the method itself, such as variants in positions with a low number of aligned sequences for evolutionary based methods such as EVE, or the lack of available scores for ensemble predictors such as REVEL. Therefore, when considering the performances of a specific tool, it is important to also report the percentage of variants that the tool can classify. The percentage of variants classified by each tool for each data set, is reported in Table 1, and the number of classified variants by each tool is shown the bar plots of Figure 2D,E. The average percentage of classified variants is 80% and 85% for the Macnee and ClinGen data sets, respectively.

3.2 | Pathogenicity prediction performances on ClinVar variants with benign/pathogenic versus likely-benign/likely-pathogenic clinical significance

We tested the performances of the prediction tools separately for the variants classified as likely benign/likely

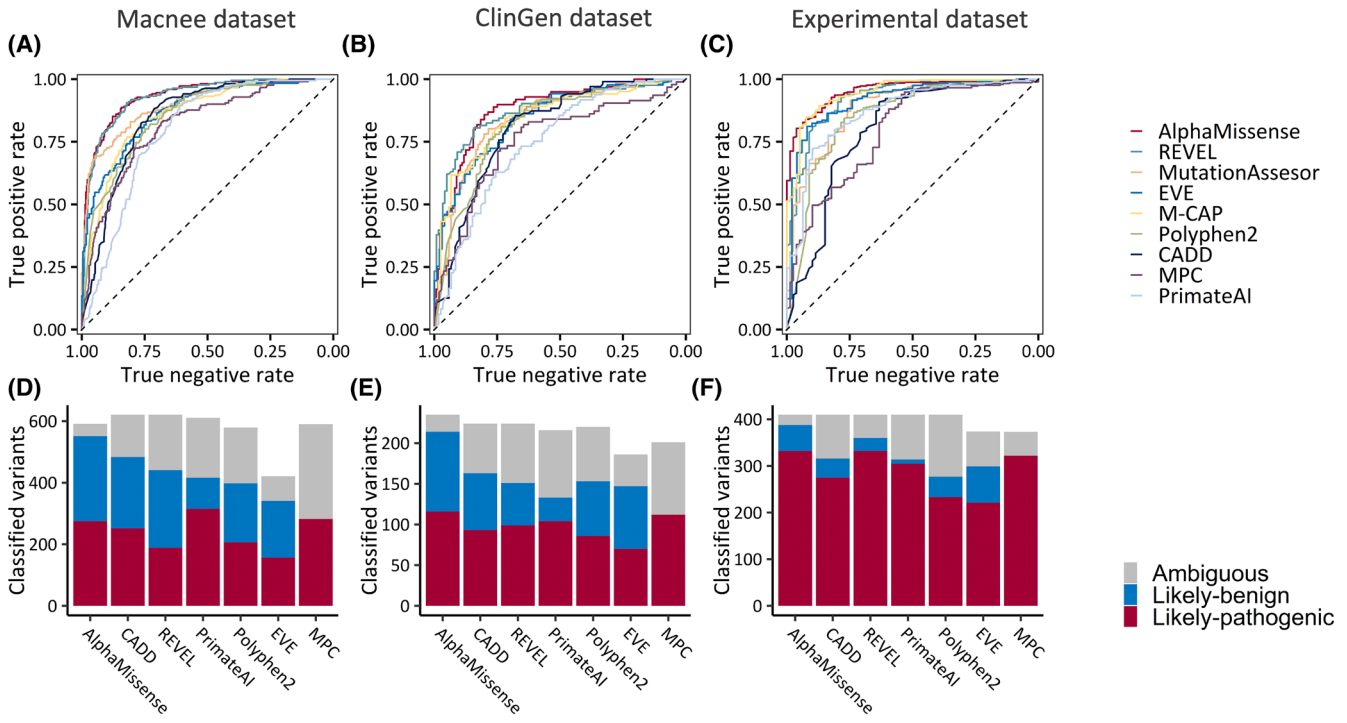

FIGURE 2 Performances and number of classified variants for each pathogenicity prediction tool on three epilepsy data sets. (A–C) ROC curves on the ClinVar Macnee, ClinVar ClinGen, and experimental data sets, respectively. (D–F) Bar plots showing the number of variants classified by each prediction tool, for the ClinVar Macnee, ClinVar ClinGen, and experimental data sets, respectively. The height of the bars corresponds to the total number of variants for which each tool provides a classification. The different colors in each bar show variants classified as pathogenic (red), benign (blue), or ambiguous (gray), respectively. MutationAssessor and M-CAP predictions are not shown in the bar plots because for these two tools no thresholds or criteria to distinguish ambiguous predictions are available (see Methods). Abbreviation: ROC curve, Receiver Operating Characteristic curve.

TABLE 1 Performance of the evaluated predictors on the two epilepsy ClinVar data sets.

	Macnee data set						ClinGen data set					
	B/P		LB/ LP		All		B/P		LB/ LP		All	
	AUC	%	AUC	%	AUC	%	AUC	%	AUC	%	AUC	%
AlphaMissense	.981	86	.917	83	.934	84	.988	93	.823	92	.880	92
Revel	.975	92	.916	86	.930	88	.959	92	.834	86	.878	88
Mutation assessor	.946	88	.885	83	.901	84	.939	87	.791	84	.841	85
EVE	.927	61	.858	59	.872	60	.963	74	.786	72	.843	73
M-CAP	.914	92	.840	86	.860	88	.926	90	.800	86	.846	88
Polyphen2	.881	89	.847	79	.855	82	.856	92	.803	84	.822	86
CADD	.872	92	.842	86	.849	88	.885	92	.761	86	.806	88
MPC	.890	83	.797	83	.816	83	.843	73	.715	82	.755	79
Primate AI	.860	90	.771	85	.795	86	.815	87	.699	84	.748	85

Note: AUC is the area under the receiver-operating characteristic (ROC)- curve, which ranges from 0 to 1 and evaluates the overall performances of a binary prediction tool across different thresholds. "%" reports the percentage of variants for which the tool is able to provide a classification. B/P is the subset of variants classified as benign or pathogenic. LB/LP is the subset of variants classifies as likely-benign or likely-pathogenic. "All" indicates all the variants of each data set.

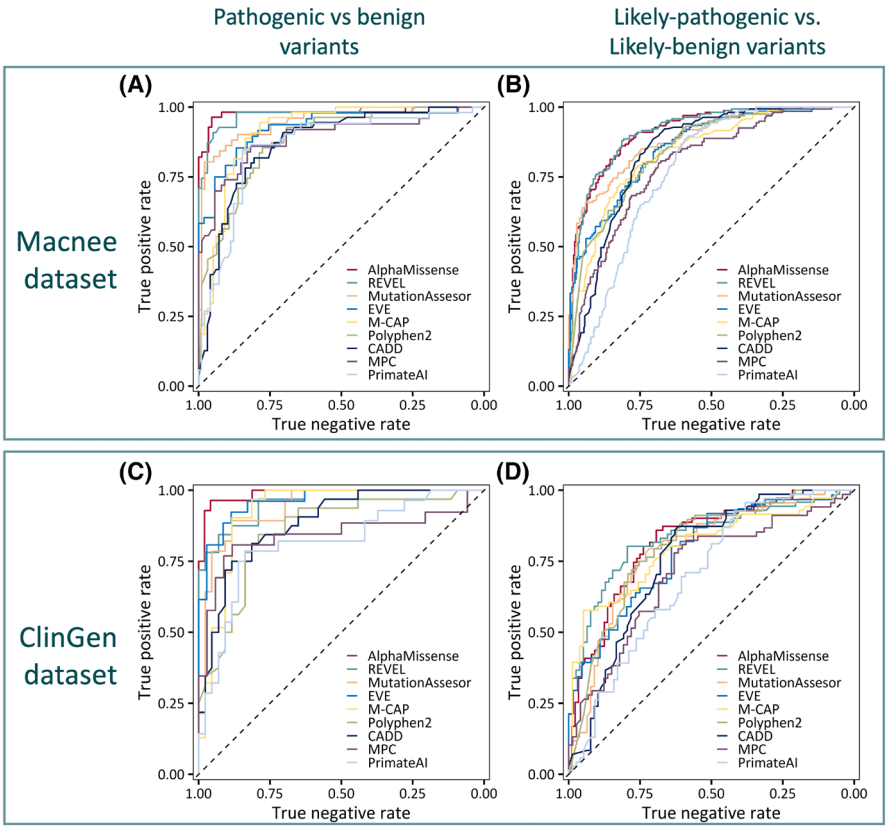

FIGURE 3 ROC curves illustrating the performances of the considered prediction tools, separately on the subset of variants classified as benign/pathogenic and likely-benign/likely/pathogenic in ClinVar for the Macnee and ClinGen-based data sets. (A) Variants classified as either benign or pathogenic in the Macnee data set. (B) Variants classified as either likely-benign or likely pathogenic in the Macnee data set. (C) Variants classified as either benign or pathogenic in the ClinGen data set. (D) Variants classified as either likely-benign or likely pathogenic in the ClinGen data set. Abbreviations: ROC curve, Receiver Operating Characteristic curve.

pathogenic (363 likely benign and 191 likely pathogenic in the Macnee data set, and 95 likely benign and 83 likely pathogenic in the ClinGen-based data set) from those classified as benign/pathogenic (106 benign and 68 pathogenic in the Macnee data set, and 48 benign and 40 pathogenic in the ClinGen-based data set). Results are reported in Table 1 and the corresponding ROC curves are shown in Figure 3. The performances on the subset the of variants classified as benign/pathogenic outperform by 6% and 16% (for the Macnee and ClinGen-based datasets, respectively) the performances derived on the data set of missense variants classified as likely-benign/likely-pathogenic.

3.3 | Pathogenicity prediction performances on variants from experimental data sets of ion channels

In the experimental data sets, the benign/pathogenic classification is the outcome of experimental characterization of mutant vs wild-type protein activity (Figure 1B). In these experimentally validated data sets, a total of 421 variants were considered (79 benign and 354 pathogenic). The overall accuracies are comparable to those found on the ClinVar-derived data sets and are shown in Figure 2C. The best performing computational tool on this data set was AlphaMissense (AUC .95) followed by REVEL (AUC .93) and EVE (AUC .92). The number of classified variants by each tool is shown in the bar plots of Figure 2F. We then sought whether there were differences in predicting variants with LoF and GoF functional effects. We, therefore, computed AUC of the considered tools separately on the subset of 86 GoF and the 79 benign, and on the subset of 155 LoF and 79 benign variants. Variants with mixed/LoF-mixed effects were excluded from this analysis. We found that all considered predictors except PrimateAI better predict for LoF than for GoF variants, with an average increase in the AUC of about 3%.

3.4 | Confidence levels in variant pathogenicity predictions

In the Results sections, we have computed the performance in pathogenicity prediction of each computational tool (estimated through the AUC, Figure 2A–C), and the percentage of variants for which each tool was able to provide a classification (number of classified variants shown in Figure 2D–F). Besides these measurements, it is also important to evaluate how many variants can be classified as benign or pathogenic with good confidence. Clearly, prediction tools that classify most variants with low confidence are of limited benefit for clinical applications. This is usually achieved by setting two thresholds on the predictions. These two thresholds separate good confidence benign, uncertain, and good confidence pathogenic predictions. Most of the prediction tools presented in this study have been calibrated previously to yield thresholds consistent with the various ACMG guideline strength-of-evidence criteria.[8] Using this calibration, we considered a high-confidence classification as one that offers at least supporting evidence regarding whether a variant is pathogenic or benign. For tools lacking a calibration according to ACMG guidelines (EVE, AlphaMissense), we derived thresholds from the original publications, which claim to provide classifications with ~90% accuracy (EVE) or 90% precision (AlphaMissense) and consider these thresholds as classifications of high confidence.

The results are shown in Figure 2D–F, where the number of variants classified with high confidence is shown in red and blue for pathogenic and benign variants, respectively, and in gray for variants classified with low confidence. Among the seven scores for which we derived thresholds, AlphaMissense was found to have the highest percentage of variants classified with high confidence with 93.2%, 91.1%, and 94.6% in the three considered data sets, respectively. It is followed by REVEL with only 71.0%, 67.4%, and 87.8% of variants classified with high confidence across the three data sets. In comparison to prior prediction tools, AlphaMissense reduced the number of variants with low to moderate confidence predictions by an average of 25.9%, and 19.2% when compared to REVEL. It is worth noting that an objective comparison when introducing thresholds is challenging because most prediction tools do not provide thresholds, and levels of confidences can be chosen arbitrarily. Although all the thresholds used for all the methods except EVE and AlphaMissense are derived within the same calibration framework,[8] the thresholds considered for the AlphaMissense predictions are not.

4 | DISCUSSION

Here, we benchmark, specifically on epilepsy-associated genes, the performance of state-of-the-art variant pathogenicity prediction tools, including the recently developed AlphaMissense and EVE. We observed that AlphaMissense and REVEL reached the best accuracies, with AlphaMissense reducing the number of VUS. Overall, these tools perform well in epilepsy-associated genes and their use is recommended when applying the ACMG/AMP guidelines. In the original iteration of the 2015 ACMG/AMP variant classification guidelines in silico algorithms could provide only supporting evidence for variant pathogenicity classification when agreement between at least three tools was reached. A recent approach that calibrated prediction scores based on a Bayesian framework, showed that some scores can achieve moderate to strong evidence and that the use of only one tool is preferable, as computational tools outcomes between different tools tend to be correlated. From this work it emerges that AlphaMissense is the most accurate pathogenicity prediction tool for epilepsy syndromes and should be the choice once calibration according to the Bayesian framework has been carried out. Calibration would introduce specific thresholds for different levels of confidence in the classification (supporting, moderate, strong, very strong) and would, therefore, prevent classification when confidence

levels are low. Indeed, although reducing the high number of VUS will have a positive impact on diagnostic yield, a VUS classification may still be preferable to a wrong classification, which may lead to incorrect therapeutic choices. Although future aggregation of expert-curated or experimental data sets will provide independent test sets to better assess the power of computational tools, the state of the art suggests that, as suggested in the 2022 update,[8] current tools can provide more than supporting evidence and can be used effectively for variant classification for epilepsy syndromes.

In summary, this study provides valuable insights by offering three possible estimations to gauge the usefulness of variant pathogenicity prediction tools in epilepsy genes within clinical settings. However, it should be noted that computational tools are aid tools that cannot replace the final clinical classification decision guided by expert opinion. Our findings inform best practices for variant classification in genetic epilepsy syndromes and emphasize the need to incorporate periodic reevaluation of genetic test results with newly developed or updated prediction tools into standard clinical practice, to improve diagnostic yield.

5 | LIMITATIONS

This comparative assessment has several important limitations to consider. A major limitation of this analysis is the absence of a set of variants with definitively known pathogenic/benign effects, as the pathogenic/benign labels collected in ClinVar may not be entirely accurate. This limitation applies to all prediction tools and comparative assessments intended to evaluate variant pathogenicity. To mitigate this issue, we excluded all variants classified with low confidence (0-star clinical review status) in ClinVar. However, most of the variants available for this study were assigned a clinical review status of 1 star ("criteria provided, single submitter"). A second limitation, known as "data leakage," is the unintentional use of information from the test set to build the model and vice versa, which can bias predictions toward inflated accuracies. To address this problem, we included only patient variants submitted to ClinVar after the development of all the evaluated tools. However, this approach may not fully eliminate the possibility that one or more of these tools might have been used in clinical label assignment of the newly added variants, thus introducing circularity in the testing process.

In contrast, the experimental data set does not suffer from these two biases, because the pathogenic/benign labels are taken from electrophysiological experiments, a gold-standard assay. Thus, clinical labels deriving from electrophysiological experiments should represent a test set with truly independent label assignments. However, other limitations apply. First, these data sets are usually small, as generating experiments is costly and time consuming. They are usually available for a small number of protein classes and tend to overrepresent patient variants compared to those found in the general population. Furthermore, electrophysiological experiments do not always capture all aspects of channel function due to non-standard protocols that may exclude more complex mechanisms such as slow inactivation, persistent current, or aberrant ionic conductance. Additional effects at the neuron and network level, or interactions with modulating proteins, may further influence the overall effect of the variant.

AUTHOR CONTRIBUTIONS

D.L. and L.M. conceived the work and wrote the manuscript. L.M. and T.B. performed the analysis. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

We thank Gary Taylor for support in setting up the analysis platform.

CONFLICT OF INTEREST STATEMENT

Neither of the authors has any conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ETHICS STATEMENT

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Ludovica Montanucci https://orcid.org/0000-0003-2375-7359
Alina Ivaniuk https://orcid.org/0000-0001-9912-3112
Samden Lhatoo https://orcid.org/0000-0002-0260-3855
Dennis Lal https://orcid.org/0000-0002-5173-9636

REFERENCES

1. Rehm HL et al. The landscape of reported VUS in multi-gene panel and genomic testing: time for a change. Genet Med off J Am Coll Med Genet. 2023;25:100947.
2. Gunning AC, Fryer V, Fasham J, Crosby AH, Ellard S, Baple EL, et al. Assessing performance of pathogenicity predictors using clinically relevant variant datasets. J Med Genet. 2021;58:547–55.
3. SoRelle JA, Pascual JM, Gotway G, Park JY. Assessment of interlaboratory variation in the interpretation of genomic test results in patients with epilepsy. JAMA Netw Open. 2020;3:e203812.

4. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–D894.
5. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.
6. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021;599:91–5.
7. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381:7492.
8. Pejaver V, Byrne AB, Feng BJ, Pagel KA, Mooney SD, Karchin R, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet. 2022;109:2163–77.
9. Macnee M, Pérez-Palma E, López-Rivera JA, Ivaniuk A, May P, Møller RS, et al. Data-driven historical characterization of epilepsy-associated genes. Eur J Paediatr Neurol EJPN off J Eur Paediatr Neurol Soc. 2023;42:82–7.
10. Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, Dwight SS, et al. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical genome resource. Am J Hum Genet. 2017;100:895–906.
11. Vasilevsky NA et al. Mondo: Unifying diseases for the world, by the world. 2022 https://doi.org/10.1101/2022.04.13.22273750
12. EpilepsyDiagnosis.org. https://www.epilepsydiagnosis.org/
13. Trinidad, M. et al. High-throughput discovery of SLC6A1 variants affecting GABA transport in neurological disorders. https://doi.org/10.1101/2022.03.09.22271804 (2022).
14. Boßelmann CM, Hedrich UBS, Müller P, Sonnenberg L, Parthasarathy S, Helbig I, et al. Predicting the functional effects of voltage-gated potassium channel missense variants with multi-task learning. EBioMedicine. 2022;81:104115.
15. Brunklaus A, Feng T, Brünger T, Perez-Palma E, Heyne H, Matthews E, et al. Gene variant effects across sodium channelopathies predict function and guide precision therapy. Brain J Neurol. 2022;awac006:4275–86. https://doi.org/10.1093/brain/awac006
16. Mermer F, Poliquin S, Rigsby K, Rastogi A, Shen W, Romero-Morales A, et al. Common molecular mechanisms of SLC6A1 variant-mediated neurodevelopmental disorders in astrocytes and neurons. Brain. 2021;144:2499–512.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
18. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011;39:e118.
19. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018;50:1161–70.
20. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48:1581–6.
21. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Editor Board Jonathan Haines Al. 2013;7:7.20.

How to cite this article: Montanucci L, Brünger T, Boßelmann CM, Ivaniuk A, Pérez-Palma E, Lhatoo S, et al. Evaluating novel in silico tools for accurate pathogenicity classification in epilepsy-associated genetic missense variants. Epilepsia. 2024;65:3655–3663. https://doi.org/10.1111/epi.18155