aiVCE – Gene-specific Artificial Intelligence-based Variant Classification Engine: Results of a Time Capsule Experiment

Yaron Einhorn, MSc[1] (yaron@genoox.com); Moshe Einhorn, MSc[1]; Adaia Kamshov, MD[1]; Oron Lev, MSc[1]; Amir Trabelsi, BSc[1]; Nurit Paz-Yaacov, PhD[1]; Susan J. Gross, MD, FACMG[1,2]

[1]Genoox, Tel Aviv, Israel; [2]Icahn School of Medicine at Mount Sinai, New York, NY, USA

ABSTRACT

Introduction: Next-generation sequencing (NGS) technology and shared archives usage has expanded, thus increasing the volume of variant classification data and the need for novel analytical approaches. Information derived from the classification of variants is critical to discovering or confirming disease etiologies and directing treatment guidelines and patient-specific plans. The complexities of variant classification, however, require assemblage and assessment of continually updated information.

We developed a novel AI-based Variant Classification Engine (aiVCE) algorithm, based on ACMG/AMP standards and guidelines, to integrate knowledge from available databases and literature and to expedite gene-specific variant classification. The aiVCE algorithm can assess all classification criteria amenable to automation and places particular emphasis on considering gene-specific evidence at the gene level, consistent with the latest efforts by ClinGen Expert Panels (EPs).

This blinded study, assessed the aiVCE's overall and rule-level performance using ClinVar variants. Evaluation of discordance between aiVCE and ClinVar uncovered evidence that may not have been available to submitting laboratories, highlighting AI's utility in variant classification. Applying AI-enhanced computational methodologies to existing guidelines for future ClinVar variant evaluation may assist the classification and interpretation of variants with limited clinical information, thus greatly reducing analytical bottlenecks.

Methods: aiVCE's overall and rule-level performance was evaluated utilizing ClinVar variants with reference/submission creation dates on or before 05/01/2017 ('Train') and after 05/01/2017 ('Test').

'Full' (75,801 variants with ≥1 star, including only submissions from CLIA-certified laboratories) and 'Increased-Certainty' (3,993 variants with ≥2 stars) datasets were evaluated.

'Test' variants were classified as pathogenic (*P*), likely-pathogenic (*LP*), uncertain significance (*VUS*), likely-benign (*LB*), or benign (*B*). *VUS* with sufficient supporting data were subclassified as *VUS*-leaning benign or *VUS*-leaning pathogenic. aiVCE results were evaluated to determine concordance with final ClinVar classification and rule-level determinations.

Results: The aiVCE demonstrated concordance among Increased-Certainty variants of >97%, while concordance was >95% across variant effects (missense, synonymous, null, splice region, intronic, untranslated region, non-frameshift indels). For the Full data set, concordance was >93.5%. When assessing the aiVCE's application of the different ACMG rules, significant differences were observed between ClinVar *P/LP* and *B/LB* variants (all p-values <0.00001) across the different rules, thus making a case for gene-specific rule selections. The aiVCE exhibited robust performance in categorizing variants as *P/LP*.

Conclusion: The aiVCE algorithm, even without input from clinical databases specific to the 'Test' set, could predict with very high concordance whether a variant would be categorized as *P/LP*. Therefore, robust algorithms that apply the latest computational methodologies to ACMG/AMP guidelines may assist variant scientists with classification and interpretation of variants, including those with limited clinical information.

Knowledge derived from computational methodologies can augment human expertise and judgment required to deduce final variant classifications which can aid clinical and research laboratory professionals in the current era characterized by increased complexity of variant analysis and interpretation.

Objective

- Validate the artificial intelligence-based Variant Classification Engine (aiVCE) using the ClinVar database through a blinded time-capsule experiment to predict the algorithm's ability to classify variants uploaded to the ClinVar database after the time capsule cutoff date.

Introduction and Methods

Artificial Intelligence-Based Variant Classification Engine (aiVCE)

- Supports the ***extraction of massive and existing evidential data*** from various data sources (databases and published literature)
  - Continual assimilation of current public variant-level information
- Is based on ***ACMG/AMP standards and guidelines plus a machine-learning process to implement variant classification rules for novel and reported*** variants
  - aiVCE automates 17/28 ACMG rules
- End users can ***input evidence for remaining rules*** (e.g., rules specific to patient genotype)
- ***Determines internal numeric Classification Score to facilitate VUS subclassification***
- ***Assigns Evidence Scores*** – for variants consolidated from ClinVar, UniProt, in-house curated evidence etc. – which create an internal *Variant* ***KnowledgeBase*** and form the basis of classifying any new variant
  - Aggregated gene-level models are generated based on the Variant KnowledgeBase to determine gene-level data e.g. frequency thresholds and known disease mechanisms for a gene
- aiVCE ***emphasizes robust sensitivity in detecting P variants***, even at the expense of specificity, as identifying all *P* variants is the focus
- ***Provides comprehensive annotation for clinical use***
- ***Benchmarked using ClinVar*** (version 2018-10) variants

Datasets

- **'Full' dataset:** 75,801 variants with ≥1 star, limited to ≥1 CLIA-certified laboratory submitter and without conflicting interpretations, spanning 3,115 different genes
- **'Increased-Certainty' dataset:** A subset dataset of 3,993 variants with ≥2 stars, across 638 different genes, variants in the Increased-Certainty dataset had ≥2 submitters with no conflicts or were EP-reviewed (https://www.ncbi.nlm.nih.gov/clinvar/docs/details/)

Legend Key: ACMG, American College of Medical Genetics and Genomics; aiVCE, artificial intelligence-based Variant Classification Engine; *B*, benign; CI, confidence interval; *LB*, likely benign; *LP*, likely pathogenic; *P*, pathogenic; *VUS*, variant of uncertain significance; *VUS-LB*, *VUS*-leaning benign; *VUS-LP*, *VUS*-leaning pathogenic

ACMG Standards and Guidelines

Criteria for classifying variants to homogenize methods and reduce discordance and ambiguity between labs.

- Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (Richards *et al.*, 2015)
- Weighted rules related to:
  - Variant frequency
  - Variant type
  - Association to previous reports for pathogenicity
  - Consistency with inheritance model
- **Evidence of pathogenicity (*P* supporting rules):** very strong (PVS1), strong (PS1-PS4, moderate (PM1-PM6), and supporting (PP1-PP5)
- **Evidence of benign impact (*B* supporting rules):** stand-alone (BA1), strong (BS1-BS4), and supporting (BP1-BP7)

Results

Variant Classification Concordance/ Discordance

- High aiVCE and ClinVar concordance with 2-tier classification based on medical importance.
  - **Increased-Certainty dataset:** 97.29% (95% CI: 96.79% – 97.79%) (**Table 1**).
  - **Full dataset:** 93.78% (95% CI: 93.61% – 93.95%) (**Table 2**).
- aiVCE can further classify *VUS* variants as *VUS-LB* or *VUS-LP* which provides useful variant prioritization.
  - Of the Full dataset (58,067 variants) 7,282 (12.5%) variants were sub-classified as either *VUS-LP* or *VUS-LB* (**Table 2**).
  - Discordance of only 1.2% variants called as *VUS-LP* by the aiVCE but were *LB* in ClinVar and only 0.05% variants called as *VUS-LB* by the aiVCE but *LP* in ClinVar.

Table 1. Benchmarking an aiVCE using a time-capsule of the ClinVar database – Increased-Certainty dataset

aiVCE / ClinVar	*B*	*LB*	*VUS-LB*	*VUS*	*VUS-LP*	*LP*	*P*
B	123	44	9	10	0	0	0
LB	24	146	63	1,286	1	1	0
VUS	0	172	41	1,327	166	49	0
LP	0	0	0	12	16	58	0
P	0	0	0	12	18	415	0

ClinVar	*aiVCE*	*B/LB/VUS*	*P/LP*	Concordance (95% CI)
B/LB		3,412	50	97.29%
P/LP		58	473	(96.79% – 97.79%)

aiVCE, artificial intelligence-based Variant Classification Engine; *B*, benign; CI, confidence interval; *LB*, likely benign; *LP*, likely pathogenic; *P*, pathogenic; *VUS*, variant of uncertain significance; *VUS-LB*, *VUS*-leaning benign; *VUS-LP*, *VUS*-leaning pathogenic

Table 2. Benchmarking an aiVCE using a time-capsule of the ClinVar database – Full dataset

aiVCE / ClinVar	*B*	*LB*	*VUS-LB*	*VUS*	*VUS-LP*	*LP*	*P*
B	2,096	1,093	203	806	9	4	0
LB	588	1,719	865	20,967	82	32	0
VUS	61	1,562	372	27,807	4,342	2,040	5
LP	5	5	3	919	1,125	4,497	3
P	5	3	1	286	280	4,007	9

ClinVar	*aiVCE*	*B/LB/VUS*	*P/LP*	Concordance (95% CI)
B/LB/VUS		62,572	2,081	93.78%
P/LP		2,632	8,516	(96.70% – 98.02%)

aiVCE, artificial intelligence-based Variant Classification Engine; *B*, benign; CI, confidence interval; *LB*, likely benign; *LP*, likely pathogenic; *P*, pathogenic; *VUS*, variant of uncertain significance; *VUS-LB*, *VUS*-leaning benign; *VUS-LP*, *VUS*-leaning pathogenic

Variants and ACMG Rule

- When considering gene-specific rules for missense variants, aiVCE differentially applied *P*-supporting (PS1, PM1, PM5, PP2, PP3) rules to *P/LP* variants, and *B*-supporting (BP1) rules to *B/LB* variants (p<0.00001 for each rule) (**Figure 1**).
  - The aiVCE differentiated between *P/LP* vs. *B/LB* missense variants.
  - Percentages were derived as number of variants for which each rule is met divided by total number of variants for each ClinVar classification (*P/LP*, *VUS*, *B/LB*).

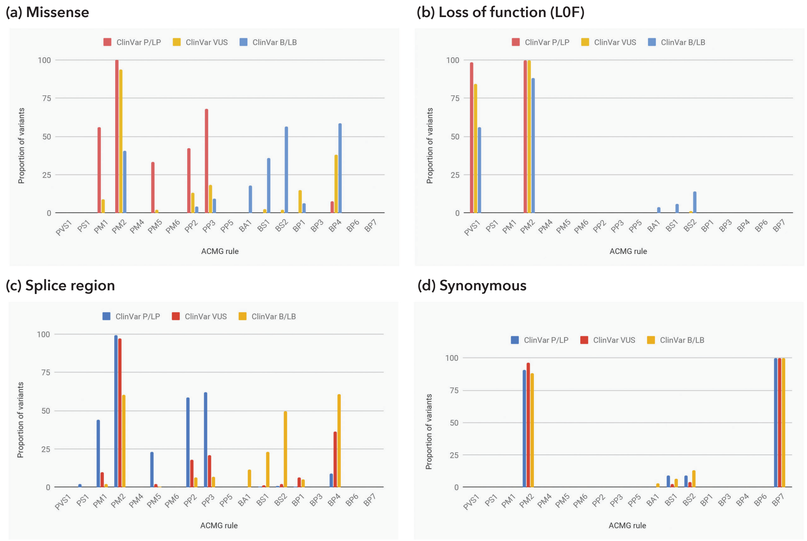

Figure 1. Distribution of ClinVar variants by aiVCE application of ACMG rules – Full dataset.

- The distribution of ClinVar variants – according to the aiVCE application of ACMG rules and ClinVar submitter classifications – demonstrated a significant difference (p<0.0001) of *P*-supporting rules, as well as for application of *B*-supporting rules, to *P/LP* vs. *B/LB* variants (**Figure 2**).

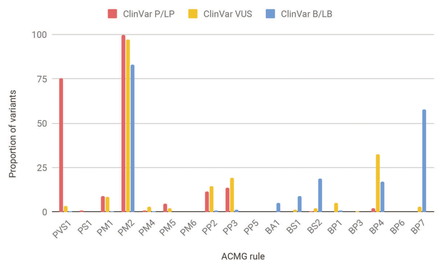

Figure 2. Distribution of ClinVar variants by aiVCE application of ACMG rules.

aiVCE Performance Across Disease Categories

- Robust performance of gene-specific frequency-related rules across various disease mechanisms.
- Employed 6 gene panels from the Genomics England PanelApp (https://panelapp.genomicsengland.co.uk/):
  - RASopathies, Hereditary Ataxia, Familial Breast Cancer, Hereditary Neuropathy, Hearing Loss, and Confirmed Fanconi Anemia (FA) or Bloom Syndrome (BS).
  - High concordance (92.67% – 98.4%) observed across the various disease categories evaluated for the Full dataset (**Table 3**).

Table 3. Concordance between the aiVCE and ClinVar classifications of *P/LP* variants by disease – Full dataset.
Based on 2-tier classification (*P/LP* vs. *B/LB/VUS*)

Disease	Number of Variants	Concordance
Hereditary Neuropathy	11,943	94.78%
RASopathies	1,194	94.52%
Hereditary Ataxia	4,553	94.24%
Familial Breast Cancer	7,164	98.25%
Hearing Loss	7,617	92.67%
Fanconi Anemia/Bloom Syndrome	3,633	98.40%

Variant Classification by Effect

- Concordance between aiVCE and ClinVar classifications showed strong agreement across all groupings even in variants typically considered difficult to classify, e.g. missense and splice region variants.
  - Missense variants comprised the largest group in both datasets (Full: 47.07%, Increased-Certainty: 47.02%), followed by synonymous variants, null variants, splice region variants, intronic and untranslated region (UTR) variants, and non-frameshift indel variants (**Table 4**, only the results for the Increased-Certainty dataset are shown).
  - Respective levels of concordance in the Full and Increased-Certainty datasets were:
    - 85.74% and 99.91% for null variants
    - 90.27% and 95.30% for missense variants
    - 99.91% and 99.92% for synonymous variants
    - 98.90% and 100% for intronic/UTR variants not located in a splice region
    - 96.58% and 97.59% for variants located in a splice region

Table 4. Distribution of variant effects overall and by ClinVar classification – Increased-Certainty dataset

Effects	All ClinVar	ClinVar *P*	ClinVar *VUS*	ClinVar *B*
Number of variants	3,993	531	1,755	1,707
Null variants, n (%)	477 (11.94%)	453	19	4
Frameshift	295	287	8	0
Nonsense	119	110	7	2
Splice Donor/Acceptor	59	56	1	2
Start-loss	4	1	3	0
Intronic and UTR, n (%)	105 (2.62%)	0	5	100
Splice Region, n (%)	291 (7.28%)	7	63	221
Synonymous, n (%)	1,204 (30.14%)	1	11	1,192
Missense, n (%)	1,878 (47.02%)	66	1,627	184
Non-frameshift Indels, n (%)	39 (0.97%)	3	30	6

CONCLUSIONS

- This study benchmarked an aiVCE algorithm, previously shown to be a robust platform for comprehensive downstream analysis. The aiVCE algorithm:
  - Demonstrates robust concordance (>97%) in determining whether future variants would be categorized as *P/LP*
  - Predicts thresholds for variant/allele frequency-based rules
  - Is sensitive and specific in classifying variants – based on observations related to gene-specific rules
- Highlights the importance of data sharing to reduce uncertainty in variant classification
  - A data-driven, AI-based tool that relies on previous evidence making data-sharing and community initiatives like ClinVar and gnomAD essential
- Development of a structured approach to incorporate additional information by databases such as ClinVar would provide evidence currently limited in aiVCE
- Examination of discordance between aiVCE and ClinVar highlights the importance of regular re-analysis
- ACMG/AMP criteria comprise more *P*-supporting than *B*-supporting rules → inclusion of additional *B*-supporting evidence may be warranted
- **Could streamline variant classification by automating ACMG rules for which supporting evidence is available**
  - Assesses all classification criteria amenable to automation
  - Provides accurate interpretation of variants for clinical purpose
  - Scales the implementation of gene-specific recommendations