PRESENTER
Navya Mrig
Grade 12, The Shri Ram School, Gurugram



A Machine Learning Framework for Predicting Speech Outcomes in Cochlear Implant Recipients Using Genetic Markers

Navya Mrig[1], Nirupma Singh[2], Dr Sumit Mrig[3]
[1]The Shri Ram School, Gurugram, Haryana [2]Dept. of Biological Sciences & Engineering, Netaji Subhas University of Technology, Dwarka, Delhi
[3]Sankalp ENT & Cochlear Implant Centre, Delhi



INTRODUCTION

Hearing loss affects **over 5%** of the world's population and is a leading cause of speech disability.

- **Cochlear implantation (CI)** is the treatment for severe–profound sensorineural hearing loss (SNHL), where hearing aids are not enough.
- Speech outcomes after CI **vary widely** — age at implantation, duration of deafness, etiology and rehabilitation all play a part.
- ~150 genes are linked to non-syndromic SNHL; **GJB2, SLC26A4, TMPRSS3** and **MYO7A** are the most frequently reported.
- **Spiral ganglion hypothesis:** variants affecting the neural auditory pathway (e.g. TMPRSS3) predict poorer CI outcomes.
- Yet genomic data are **rarely built into predictive models** of CI success.

OBJECTIVES

Build an **AI model** that estimates CI success **before surgery** from gene variants and demographic data.

› Identify **hub genes and pathways** through gene-interaction network analysis.
› Predict both **CI success (yes/no)** and the **post-operative speech score**.
› **Validate** the model on unseen patient data.

DISCUSSION

- **Gene and age** were chosen as key predictors for their biological role in auditory performance.
- **Random Forest** handles mixed categorical and numeric data reliably; two models give a yes/no answer and a quantitative score.
- Favourable outcomes with **CDH23, SLC26A4, GJB2** and **MYO15A** agree with published genotype–phenotype da
- **MYO7A** predicted poorer outcomes, unlike earlier reports of good CI results in Usher syndrome.
- **Earlier implantation** predicted better speech, confirming age as a key prognostic factor.
- **Limitations:** moderate dataset; rare variants (EYA1, MYO7A) under-represented; duration of deafness, residual hearing, comorbidities and rehabilitation not modelled.

MATERIALS & METHODS

1 Data collection
Three published datasets: WFS1 cohort (genotype + CI word/sentence scores), TMPRSS3 cohort (n = 127) and a cochlear transcriptomic dataset (56 DEGs).

2 Data processing
Cleaning and standardisation; age ranges converted to numeric averages. Missing age/speech scores imputed by gene & source; gene/implant type by cohort mode.

3 Network & pathway analysis
Gene-interaction network from STRING, visualised in Cytoscape. Hub genes ranked by degree, betweenness, closeness & clustering coefficient. Pathways from Reactome.

4 Model building
Python Random Forest models using gene + age: a classifier for CI success (speech score > 0) and a regressor for the speech score. 80 : 20 train–test split.

5 Validation
Both models applied to 13 unseen patients (Sankalp ENT & Cochlear Implant Centre) to predict success and speech score.

RESULTS: GENE NETWORK

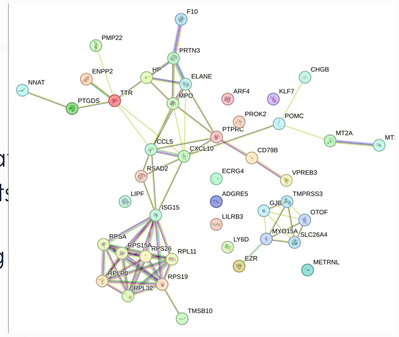


Figure 1.
STRING gene-interaction network: **36 nodes, 75 edges**. Each node is a protein; colours mark functional clusters; lines are protein–protein interactions. **ISG15** sits at the centre, linking the ribosomal cluster to the immune genes.

RESULTS (CONTINUED)

Top 10 hub genes (Table 1)

Gene	Degree	Betweenness	Closeness
ISG15	10	0.423	0.452
CXCL10	8	0.455	0.528
RPS19	8	0.071	0.346
CCL5	7	0.201	0.491
RPS15A	7	0	0.341
RPS26	7	0	0.341
RPSA	7	0	0.341
RPL32	7	0	0.341
RPLP0	7	0	0.341
RPL11	7	0	0.341

Top 5 Reactome pathways

- Influenza infection (R-HSA-168255)
- Peptide chain elongation (R-HSA-156902)
- PELO:HBS1L & ABCE1 ribosome dissociation on non-stop mRNA (R-HSA-9954714)
- Nonsense-mediated decay independent of EJC (R-HSA-975956)
- Eukaryotic translation elongation (R-HSA-156842)

8 of the 10 hub genes (ISG15 + 7 ribosomal genes) fall within the top 25 pathways.

Random Forest model performance

94.4%	0.978	0.0049
Classifier accuracy	Regressor R^2	Mean squared error

Classifier	Precision	Recall
Class 0 – unsuccessful	100%	75%
Class 1 – successful	93%	100%

Validation (13 unseen patients): predictions followed known biology — **younger age** and certain gene variants gave **higher predicted speech scores**.

Data completeness: age 100% imputed (n = 162), gene type ~69%, implant type ~40%.

CONCLUSION

Combining **gene variants with age**, the Random Forest model predicts CI speech outcomes with **high accuracy**.

- A precise tool for **personalized pre-operative counselling** of parents on expected speech & language outcomes.
- Helps **prioritize early intervention** for children most likely to benefit.
- Links **molecular pathways** with functional rehabilitation and speech outcomes.

FUTURE DIRECTIONS
Larger, diverse cohorts · rare variants · imaging & auditory-training predictors · longitudinal modelling of post-implant progress

REFERENCES

Sappington JM. Hearing loss: the silent pandemic that is finally being heard. Mo Med. 2024;121(5):350-4.
1. Eshraghi AA, et al. Genotype-phenotype correlation for predicting cochlear implant outcome. Front Genet. 2020;11:678.
2. Carlson RJ, et al. Genetic diagnoses for childhood-onset hearing loss and CI outcomes. JAMA Otolaryngol Head Neck Surg. 2023;149(3):212-22.
3. Nishio SY, Usami SI. Outcomes of CI for patients with specific genetic etiologies. Acta Otolaryngol. 2017;137(7):730-42.
4. Eppsteiner RW, et al. Prediction of CI performance by genetic mutation: the spiral ganglion hypothesis. Hear Res. 2012;292:51-8.
5. Shearer AE, et al. Genetic variants in the peripheral auditory system affect adult CI performance. Hear Res. 2017;348:138-42.
6. Han JH, et al. Comprehensive prediction model, including genetic testing, for CI outcomes. Ear Hear. 2023;44(1):223-31.
7. Usami SI, et al. Cochlear implantation from the perspective of genetic background. Anat Rec. 2020;303(3):563-93.

ACKNOWLEDGEMENTS

We thank **Aashna Saraf** for valuable feedback and guidance, and **Sankalp ENT & Cochlear Implant Centre, Delhi**, for providing the validation data.