

1 **Title:** Machine Learning classifier built with heavy metal signature biomarker genes as
2 features to distinguish between heavy metal exposure from non-heavy metal exposure
3 gene expression samples.

4 **NAMES AND AFFILIATIONS OF EACH AUTHOR**

5 Shradha Mukherjee, PhD

6 Independent Researcher.

7 This work is self-owned and not affiliated with any organization/university.

8 **CORRESPONDING AUTHOR'S NAME AND EMAIL ADDRESS**

9 Address correspondence to: Shradha Mukherjee; Independent Researcher, Calcutta,

10 West Bengal, India

11 Email smukher2@gmail.com

12 **FINANCIAL INTERESTS OR CONFLICT OF INTEREST (IF APPLICABLE)**

13 None

14 **FUNDING**

15 None

16 **Acknowledgements:**

17 Author affiliation: None. Author is an independent researcher without any affiliation.

18 Author contributions: S.M. ideated and managed the work presented in this paper. S.M.
19 also performed experimental design, wrote the manuscript, prepared the figures/tables
20 and submitted the manuscript. SM is also the corresponding author.

21 Funding: None

22 Publication: The author is most grateful to HAL (Hyper Articles en Ligne) Open Science
23 repository at France for preserving and hosting this publication with citation link for free.

24 Research Site: This work was conducted at the author's home using personal resources
25 at Sarat Ghosh Garden Road, Calcutta, West Bengal, India 700031 in early 2022 and
26 entire 2023.

27 Computational and Manuscript preparation platform: The computation was performed at
28 the author's home in Calcutta, India, mentioned above, on personal laptops (MacBook
29 Pro and Ubuntu OS) with low-speed interrupted internet connections from various
30 available providers.

31 Full text publications access: The author is most grateful to NCBI NLM NIH U.S.
32 (National Center for Biotechnology Information, National Library of Medicine, National
33 Institute of Health) www.ncbi.nlm.nih.gov/pmc/ PubMed Central and EMBL-EBI E.U.,
34 U.K. (European Molecular Biology Laboratory, European Bioinformatics Institute)

47 <https://europepmc.org> Europe PubMed Central for open-access to full-text scientific
48 publications.

49

50 Datasets: The author is grateful to the authors who submitted their datasets
51 GSE166297, GSE128431, GSE63935 and GSE126786 (cited in methods and
52 references) to NCBI NLM NIH GEO U.S. (Gene Expression Omnibus) repository
53 <https://www.ncbi.nlm.nih.gov/geo>, and to the latter for hosting and mapping these RNA-
54 seq datasets.

55

56 Computational pipelines: Citations and descriptions of all computational pipelines have
57 been given under methods sections and references. The author is grateful for these
58 computational tools.