RESEARCH ARTICLE

Editorial Process: Submission:04/21/2025 Acceptance:11/18/2025 Published:11/21/2025

Mitochondrial Epigenetic Alterations Induced by Biomass Smoke Exposure and Their Role in Age-Related Breast Cancer Risk: A Machine Learning-Based Predictive Study

Prasan Kaur[1], Nikita Soni[1], Vikas Gurjar[1,2], Nazim Nazeer[1,3], Apoorva Chouksey[1], Rajnarayan Tiwari[1], Aneha K Rajan[1], Pooja Ratre[1], Arpit Bhargava[1,3], Rupesh K. Srivasatava[4], Pradyumna Kumar Mishra[1,2]*

Abstract

Objective: To investigate the relationship between age, mitochondrial epigenetics, and BC risk among women exposed to biomass smoke, and the development of a predictive model for BC detection. **Methods:** A cross-sectional study was conducted among a total of 205 women exposed to biomass smoke and were divided into two age groups (18-25 and >25 years). mtDNA methylation, inflammatory cytokines (IL-6, TNF-α, IL-10), and carcinoembryonic antigen (CEA) levels were assessed. Machine learning models were developed using clinical and molecular data to predict BC risk. **Results:** Prolonged HAP exposure was assoc to increased mitochondrial dysfunction, particularly in older women. mtDNA methylation changes were significantly correlated with elevated CEA levels, signifies a role in BC risk. Multivariate analysis revealed strong positive correlations between age and inflammatory cytokines: IL-6 ($R = 0.95$, $p < 0.001$), TNF-α ($R = 0.99$, $p < 0.000$), and IL-10 ($R = 0.88$, $p < 0.005$), indicating heightened inflammation with age. Logistic Regression outperform predictive performance with accuracy: 90.18% and AUC: 1.00. **Conclusion:** Age and mitochondrial epigenetic changes such as mtDNA methylation and inflammatory cytokine levels are strongly linked to BC risk in women exposed to biomass smoke. These results highlight the role of mitochondrial epigenetics in BC and the potential of AI-based tools for early detection in high-risk populations. However, the study's cross-sectional design limits causal inference, emphasizing the need for longitudinal studies to clarify timing and causality.

Keywords: Particulate Matter- Air Pollution- Indoor Air Pollution- Environmental Epigenetics- Machine Learning

Asian Pac J Cancer Prev, **26 (11)**, 4069-4078

Introduction

Breast cancer (BC) is a significant global health issue, particularly among women, with 2.3 million diagnoses and 670,000 deaths in 2022 [1], representing a quarter of all female cancers. In South-East Asia, BC deaths are projected to rise by 61.7% by 2040, and in India, it accounts for 28.2% of female cancers, with an estimated 98,337 deaths in 2022 [2, 3]. Approximately 2.4 billion people in rural areas rely on solid fuels for cooking [4, 5], leading to harmful emissions such as carbon monoxide (CO) and Particulate Matter (PM), which are linked to adverse health effects in women. The average concentration of black carbon was 14.54, 14.28, and 24.69 $\mu g/m^3$ in various kitchen styles, including indoor, outdoor, and semi-open kitchens. Notably, it was also estimated that the excess lifetime carcinogenic risk in these kitchens with cooking for around four hours a day was 1.25×10^{-7}, 1.22×10^{-7}, and 2.12×10^{-7}, respectively [6]. An increase in Particulate Matter with a diameter of 2.5 micrometers or smaller (PM2.5) levels has been associated with higher odds of BC [7, 8]. Our previous research indicates that indoor air pollution may cause mitochondrial stress and alter epigenetic mechanisms [9, 10], which could increase BC risk. The Surveillance, Epidemiology, and End Results (SEER) database shows that the probability of developing BC rises with age, from 2.4% at ages 50-59 to 7.0% at 70 and older [11, 12]. This study analyzed peripheral lymphocyte expression profiles to assess their role in BC among women exposed to household air pollution (HAP) and aimed to determine the prognostic value of these profiles. Our transcriptome analysis indicated

[1]*Division of Environmental Biotechnology, Genetics & Molecular Biology (EBGMB), ICMR-National Institute for Research in Environmental Health (NIREH), Bhopal, India.* [2]*Faculty of Medical Research, Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India.* [3]*Faculty of Science, Ram Krishna Dharmarth Foundation (RKDF) University, Bhopal, India.* [4]*Department of Biotechnology, All India Institute of Medical Sciences (AIIMS), New Delhi, India. *For Correspondence: pkm_8bh@yahoo.co.uk. Prasan Kaur, Nikita Soni, Vikas Gurjar and Nazim Nazeer have equal contribution in this study.Bus*

regulation of numerous genes associated with BC progression [19], highlighting the need to understand the impacts of indoor air pollutants for better risk assessment strategies. Alongside, Artificial intelligence (AI)-driven approaches have the potential to enhance early detection and personalized treatment strategies for BC significantly. Collectively, this pilot research explores the potential interplay between biomass smoke exposure, mitochondrial epigenetic modifications, and age-associated BC risk, establishing a conceptual and methodological basis for future in-depth research. These findings contribute to a growing body of evidence linking environmental exposures to molecular alterations and provide a critical foundation for large-scale validation studies. Eventually, this work aims to inform the development of personalized cancer prevention strategies tailored to environmentally exposed populations.

Study design and sample collection

Women aged 18 to 70 (n=205) from rural areas of Morena, Sagar, Chhindwara, Gwalior, Dhar, and Betul in Madhya Pradesh, India, were recruited for the study, focusing on those regularly exposed to biomass smoke (Figure S1). Participants were divided into two groups: 23 individuals in Group A (ages 18-25) and 182 in Group B (over 25) to examine differences in expression biomarkers. Inclusion criteria included healthy homemakers and non-smokers who used biomass for cooking. Pregnant or lactating women, smokers, and those with chronic illnesses were excluded. The study was approved by the Institutional Ethics Committee, following ICMR guidelines, and informed consent was obtained from all subjects. A questionnaire collected data on socioeconomic status, habits, health, and exposure (Table S1).

Materials and Methods

Peripheral blood samples were collected via venipuncture, followed by plasma separation and lymphocyte isolation using density gradient centrifugation. DNA, RNA, miRNA, and protein were extracted from the isolated lymphocytes for downstream analyses. Oxidative DNA damage was assessed using Formamidopyrimidine Glycosylase (FPG) digestion, and Reactive oxygen species (ROS) levels were measured with 5-(-6)-Chloromethyl-20, 70-dichlorodihydrofluorescein Diacetate Acetyl Ester (CM-H_2DCFDA) [13]. Mitochondrial dysfunction was evaluated by examining the expression of the following; fission/fusion, Integrated Stress Response (ISR), and repair genes, along with mitochondrial DNA methylation[14]. To assess inflammation, levels of Nuclear Factor kappa B (NF-κB), Interleukin-6 (IL-6), and Tumor Necrosis Factor Alpha (TNF-α) were analyzed, while breast cancer biomarkers Granulocyte-Colony Stimulating Factor (G-CSF) , Alpha-fetoprotein (AFP), CEA were measured with Enzyme Linked Immunosorbent Assay (ELISA) [15]. The expression of miRNAs ;miR-7a, miR-7b, miR-21 was profiled using poly-A tailing, cDNA synthesis, and RT-PCR (Insta Q-96, India) [16]. Expression levels of lncRNA/mRNA targets [17, 18], RNA methylation, and Telomerase Reverse Transcriptase (TERT) were also analyzed [18-20]. Data were processed using R, and multivariate correlation analysis was carried out using machine learning techniques. Building on our earlier work in predictive modeling [21-23], we developed an AI-based breast cancer risk prediction model. This model integrates clinical and molecular biomarker data, with logistic regression selected for its interpretability and reliable performance. For detailed methodology please refer to Supplementary Information: Materials and Methods.

Results

Estimation of household air pollutant exposure assessment

Pearson correlation analysis was conducted to examine the relationships among PM, age (AG), and total exposure (TE) (Figure S2). The correlation matrix heatmap revealed a very weak negative correlation between age and particulate matter (AG vs PM: -0.04), indicating minimal impact of age on PM exposure. Subsequently, a slight positive correlation was observed between age and total exposure (AG vs TE: 0.53), suggesting that total exposure rises with age. A more pronounced positive correlation was established between PM levels and total exposure (PM vs TE: 0.66), indicates greater PM levels are associated with increase in total exposure. The scatter plot matrix and line plots reinforced these patterns visually, indicating no relationship between AG and PM, a moderate upward trend between AG and TE, and a strong upward trend between PM and TE. The bar chart illustrating the correlation coefficients further demonstrated these associations, with AG vs PM close to zero, AG vs TE moderate, and PM vs TE strong.

Multivariate correlation analysis

Age and inflammation levels

The immunological response to aging is significantly correlated with cytokine concentrations by age. Significantly, IL-6 levels rise with age ($R = 0.95$, $p < 0.001$), which exacerbates inflammation, especially in women who have been exposed to Indoor Air Pollution (IAP). Furthermore, TNF-α and IL-6 have a substantial association ($R = 0.99$, $p < 0.000$), indicating a coordinated inflammatory response. Age also has a significant correlation with the anti-inflammatory cytokine IL-10 ($R = 0.88$, $p < 0.005$), indicating an adaptive mechanism to combat inflammation. Additionally, IL-4 levels increase with age ($R = 0.86$, $p < 0.001$). The intricate interaction of inflammatory and anti-inflammatory signals is indicated by the significant connection between TNF-α and IL-10 ($R = 0.95$, $p < 0.000$) (Figure S3).

Analysis of mitochondrial epigenetic alterations

Variables such as AGE, oxidative damage, Repair genes, the ISR gene, mitochondrial biogenesis factors (fusion and fission), TFAM, and mitochondrial DNA (mtDNA) methylation were analyzed. Each plot includes a line of best fit with R-squared (R^2) values ranging from 0.81 to 0.97, indicating strong correlations, and p-values of 0.000, indicating statistical significance. This illustrates the relationships between aging and various

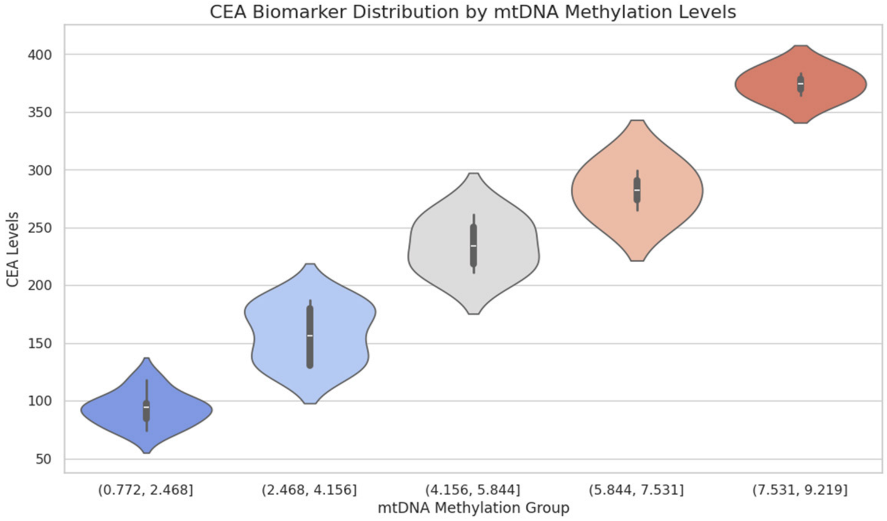
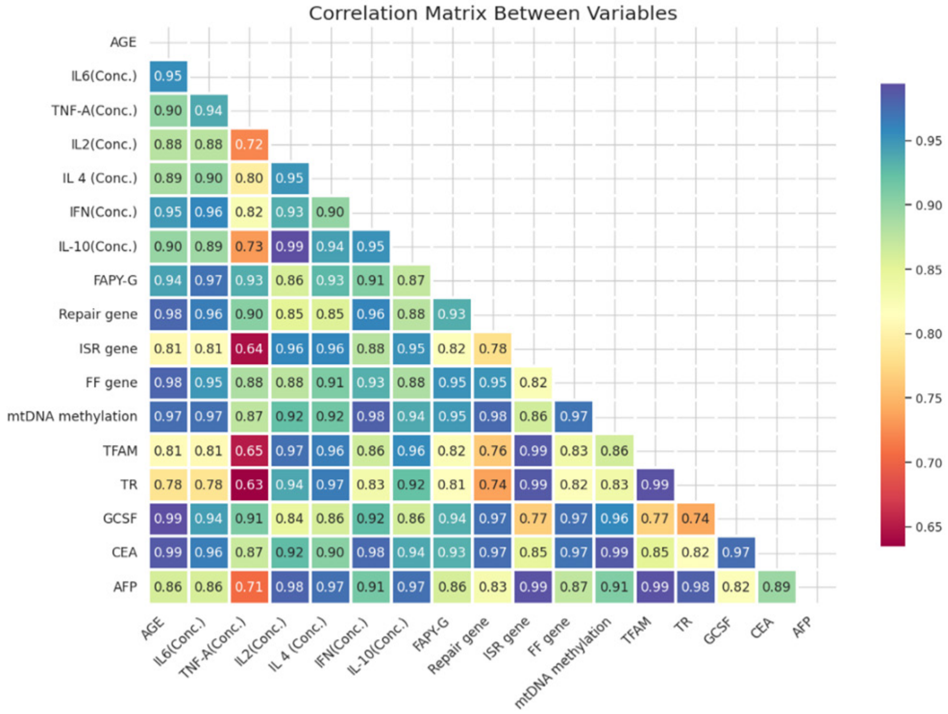


Figure 1. This Type of Plot Combines Features of Both a Box Plot and a Density Plot to Illustrate the Distribution and Density of Data Points. The x-axis categorizes different methylation levels, while the y-axis displays CEA levels ranging from 0 to 400. This distribution suggests potential correlations, indicating that higher mtDNA methylation may be linked to increased CEA levels. The plot visually conveys how CEA levels are distributed across varying methylation levels, with each group showing how these levels fluctuate. The pattern suggests a potential correlation, where higher mtDNA methylation levels may be associated with elevated CEA levels in BC cases, offering insight into possible biological relationships between these factors.

epigenetic factors (Figure S4). We emphasize the disease's heterogeneity and suggest that mtDNA methylation may influence tumor behavior and exposed individual outcomes (Figure 1). A correlation matrix displays the relationships between BC-related genes or biomarkers (Figure 2). In contrast, a scatterplot matrix visualizes the correlations among variables such as Mitochondrial Transcription Factor A (TFAM), CEA, G-CSF, AFP, and mtDNA methylation (Figure 3). A positive correlation was found, indicating that IL-6 levels rise with age, which is associated with chronic inflammation. Additionally, higher mtDNA methylation levels correlate with increased TNFα


Correlation Matrix Between Variables

	AGE	IL6(Conc.)	TNF-A(Conc.)	IL2(Conc.)	IL 4 (Conc.)	IFN(Conc.)	IL-10(Conc.)	FAPY-G	Repair gene	ISR gene	FF gene	mtDNA methylation	TFAM	TR	GCSF	CEA	AFP
AGE																	
IL6(Conc.)	0.95																
TNF-A(Conc.)	0.90	0.94															
IL2(Conc.)	0.88	0.88	0.72														
IL 4 (Conc.)	0.89	0.90	0.80	0.95													
IFN(Conc.)	0.95	0.96	0.82	0.93	0.90												
IL-10(Conc.)	0.90	0.89	0.73	0.99	0.94	0.95											
FAPY-G	0.94	0.97	0.93	0.86	0.93	0.91	0.87										
Repair gene	0.98	0.96	0.90	0.85	0.85	0.96	0.88	0.93									
ISR gene	0.81	0.81	0.64	0.96	0.96	0.88	0.95	0.82	0.78								
FF gene	0.98	0.95	0.88	0.88	0.91	0.93	0.88	0.95	0.95	0.82							
mtDNA methylation	0.97	0.97	0.87	0.92	0.92	0.98	0.94	0.95	0.98	0.86	0.97						
TFAM	0.81	0.81	0.65	0.97	0.96	0.86	0.96	0.82	0.76	0.99	0.83	0.86					
TR	0.78	0.78	0.63	0.94	0.97	0.83	0.92	0.81	0.74	0.99	0.82	0.83	0.99				
GCSF	0.99	0.94	0.91	0.84	0.86	0.92	0.86	0.94	0.97	0.77	0.97	0.96	0.77	0.74			
CEA	0.99	0.96	0.87	0.92	0.90	0.98	0.94	0.93	0.97	0.85	0.97	0.99	0.85	0.82	0.97		
AFP	0.86	0.86	0.71	0.98	0.97	0.91	0.97	0.86	0.83	0.99	0.87	0.91	0.99	0.98	0.82	0.89	

Figure 2. The Heatmap Conveys the Relationships between the Selected Variables Listed with Age through a Color-Coded System. Red squares indicate a strong positive correlation, where co-regulated gene groups or biological pathways increase together. In contrast, blue squares represent a strong negative correlation, showing that as one variable rises, the other declines. Yellow squares highlight weak or no correlation, suggesting minimal or absent relationships. This visual representation aids in identifying patterns of co-regulation and interactions among biological pathways.

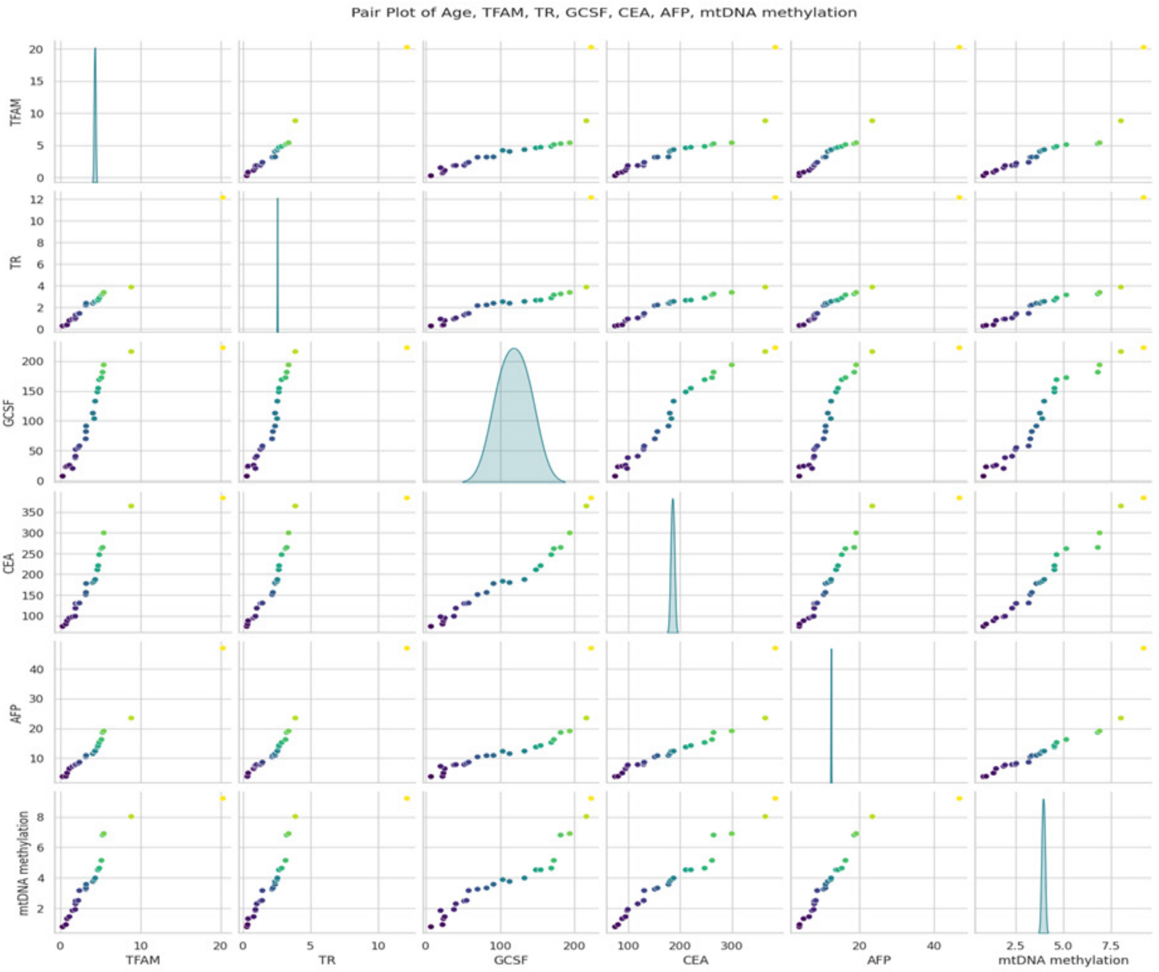


Figure 3. The Scatter Plots and Pair Plots are Color-Coded to Represent Different Groups, such as Various Stages of BC or Distinctions between Control and High-Risk Group Groups. Along the diagonal of the matrix, histograms display the distribution of each individual variable. The histogram for TFAM reveals its frequency present in the dataset in correlation with the CEA, GCSF, AFP, TR (TERT), and mtDNA methylation. This combined approach facilitates a clear visualization of relationships between variables while illustrating individual measurements' distribution.

concentration (Figure 4). Notably, older women exhibited a higher rate of RNA methylation (Figure S5).

Our results showed altered expression of miRNA-21 with fold change of 2.9255±0.05, highest in the women,

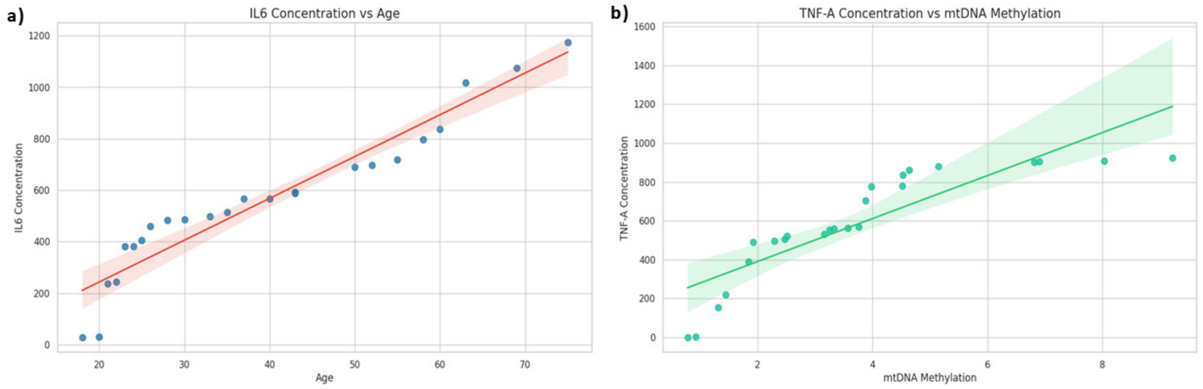


Figure 4. Features Scatter Plots that Provide a Comprehensive Overview of the Relationships Depicted in the Plot. The X-axis represents a) Age and b) mtDNA methylation, showcasing the age of individuals and revealing age-related trends. The Y-axis displays IL-6 and TNFα concentration, indicating potential inflammatory responses. Each data point corresponds to an individual and is color-coded to distinguish between different groups, such as control and High-Risk Group. A trend line illustrates the positive correlation between age and IL-6 concentration, indicating that IL-6 levels also tend to rise as age increases. Additionally, higher levels of mtDNA methylation are linked to increased TNFα concentration, highlighting a connection between genetic regulation and inflammatory responses. These findings underscore the relevance of age, IL-6, and TNFα in understanding chronic inflammation and its health implications, emphasizing the importance of monitoring these factors with aging and inflammation.

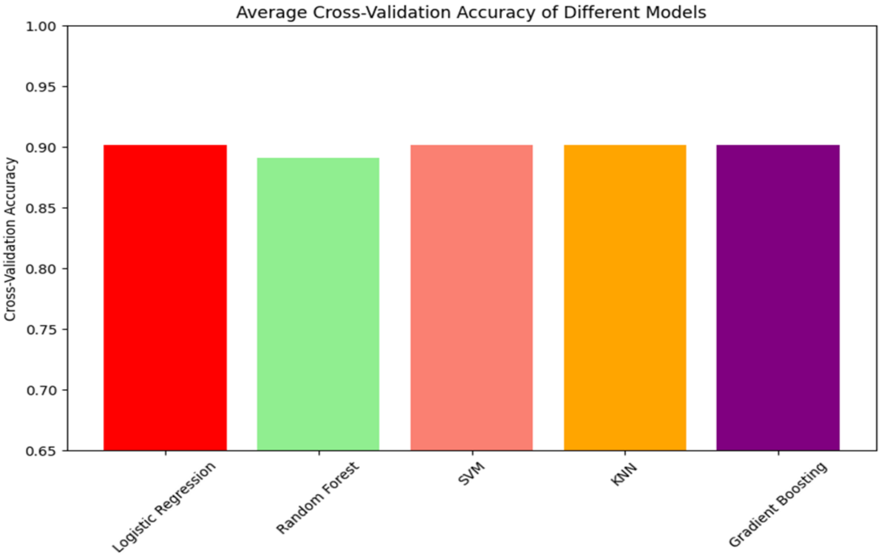


Figure 5. The Bar Chart Illustrates the Average Cross-Validation Accuracy of Five Machine Learning Models Applied to a Dataset, Potentially Related to BC Classification or Prediction. Logistic Regression shows the highest cross-validation accuracy, just above 0.90, making it the best-performing model in this comparison. Gradient Boosting follows closely, with an accuracy slightly below 0.90, while SVM and KNN perform similarly, around 0.88. Random Forest, though slightly lower in performance, achieves an accuracy of approximately 0.87. The Y-axis ranges from 0.65 to 1.0, where all models demonstrate high accuracy, indicating their effectiveness in BC prediction. Overall, this graph suggests that while all the models are reliable, Logistic Regression slightly outperforms the others, making it the most accurate for this dataset.

in group B among all miR-7a, miR-21, and miRNA-7b studied (Figure S6 a). Our results showed altered expression of lncRNAs, i.e., Gas-5, H19, and PVT, among the women exposed to HAP. The PVT was highly expressed with a fold change of 2.61755±0.05 in the Group B women category and found to be downregulated in Group A women (Figure S6 b). The expression of PTEN, MYC, PIK3CA and BCL2 were found to be downregulated and the expression of PDCD4, FOXO3, APAF1, CDK2, E2F2 and CCND1 were highly upregulated with fold change 2.05198±0.005, 1.15871±0.05, 2.08744±0.05, 1.08706±0.005, 3.44449±0.05 respectively, in Group B females in comparison to Group A women (Figure S6 c).

AI-based BC risk prediction models were developed using a dataset of genetic expression and clinical records, which included features like age, mitochondrial DNA methylation, G-CSF, and TERT levels. Various machine learning algorithms were applied, such as Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting (Figure 5). Stratified K-Fold cross-validation yielded average accuracy scores: Logistic Regression (90.18%), Random Forest (89.14%), SVM (90.18%), KNN (90.18%), and Gradient Boosting (90.18%). Logistic Regression achieved the highest accuracy with a confusion matrix showing 29 true negatives and 10 true positives, resulting in 100% recall and precision for both classes. The F1-score was 1.00, and the Receiver Operating Characteristic (ROC) curve showed an Area under curve (AUC) of 1.00, indicating excellent discrimination (Figure 6). The Precision-Recall curve also confirmed high precision and recall rates. The trained Logistic Regression model is saved for future use in clinical settings for timely BC risk assessments (Figure 7).

Discussion

In this study, we explored the impact of PM and black carbon exposure on age-related changes in mitochondrial epigenetics, which may increase BC susceptibility. We conducted a cross-sectional pilot study in India with two groups: 23 younger adults (Group A, 18-25) and 182 older individuals (Group B, above 25). Results showed significant differences in mitoepigenetic profiling in Group B, indicating that aging-related mitochondrial changes may increase BC susceptibility. We investigated mtDNA methylation status and found alterations linked to aging and environmental factors, with the older group exhibiting higher DNA expression levels. We noted that PM exposure correlates with abnormal DNA methylation patterns. Our result indicate elevated levels of DNA repair enzymes OGG1, APE1, and POLG in the older group, confirming base excision repair (BER) activation in response to oxidative damage. Notably, OGG1 activity was significantly higher than APE1's, suggesting that APE1 enhances OGG1 function. The relationship between OGG1 and unhealthy aging, especially with increased TNF-α levels, emphasizes the role of mitochondrial dysfunction in cancer susceptibility [24]. miRNAs crucial in regulating mitochondrial stress-related gene transcription and act as a link between genomic and epigenomic mechanisms [25]. Our study found altered miRNA expression in women exposed to PM2.5 and black carbon, with older women showing significantly increased miR-21, which has both tumor-suppressive and oncogenic roles. Previous research also linked PM2.5 exposure to miRNA expression changes [26]. lncRNAs influence angiogenesis, mitochondrial function, and mRNA regulation via miRNA interaction

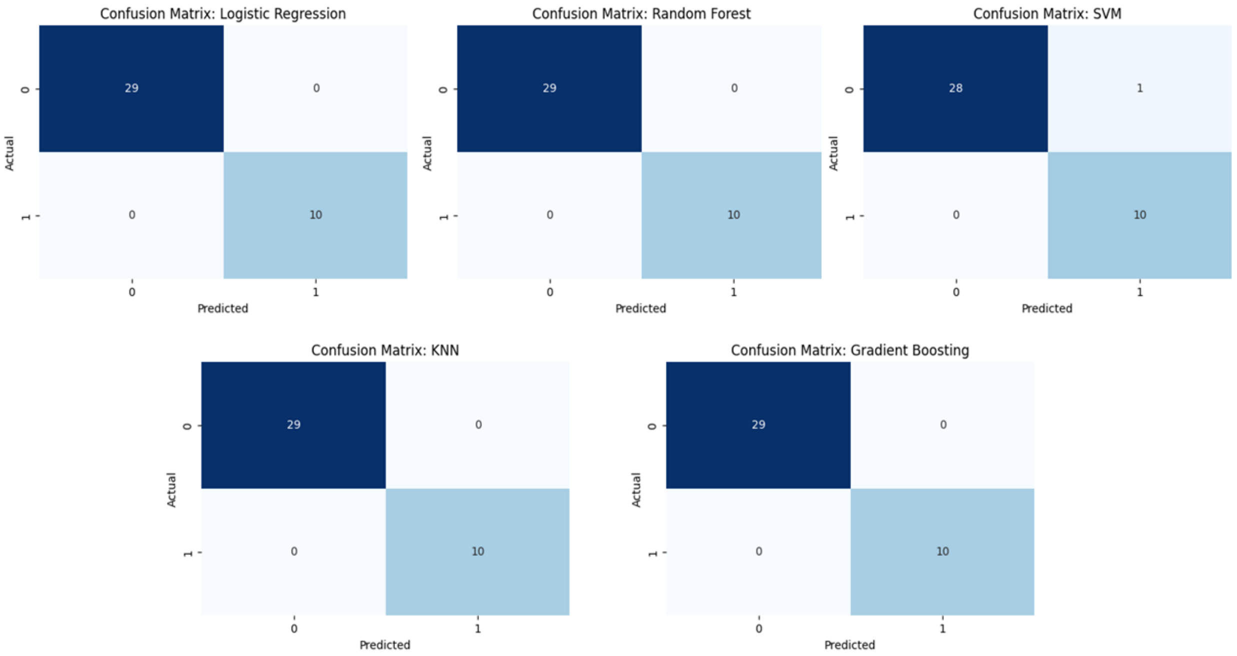


Figure 6. The Confusion Matrices for the Five Machine Learning Models Used to Predict BC are Represented as Follows: each corresponds to a different algorithm. a) The Logistic Regression model illustrates its performance in classifying cancer cases. b) The Random Forest model highlights its ability to balance sensitivity and specificity. c) The SVM emphasizes its strength in identifying true negatives while showing a tendency for false positives. d) The KNN model showcases effective true positive and true negative rates. Finally, the e) Gradient Boosting model reveals challenges with misclassifications. These matrices provide valuable insights into the strengths and weaknesses of each algorithm in predicting BC.

[27]. Increased PVT in older women and downregulation of PVT and H19 in Group A indicate ncRNA imbalance, potentially promoting mitochondrial dysfunction and tumor progression. Our findings revealed notable change in mRNA expression associated to stress and proliferation-related genes, especially in Group B, suggesting disruption of the miRNA-lncRNA-mRNA axis due to BC exposure. Additionally, we found age-related changes in the expression of OMA1, DELE1, and HRI genes in Group B, with Previous research has found that abnormalities in mitochondrial integrity are closely linked to ISR induction [16]. Along with increased proinflammatory cytokines, reflecting chronic inflammation linked to aging. Following this Studies have linked aging to chronic "inflamm-aging" driven by immunosenescence and adipose-derived cytokines [28]. Elevated IL-6, IL-8, and TNF-α may promote tumor growth [29]. This study suggests a connection between elevated cytokine

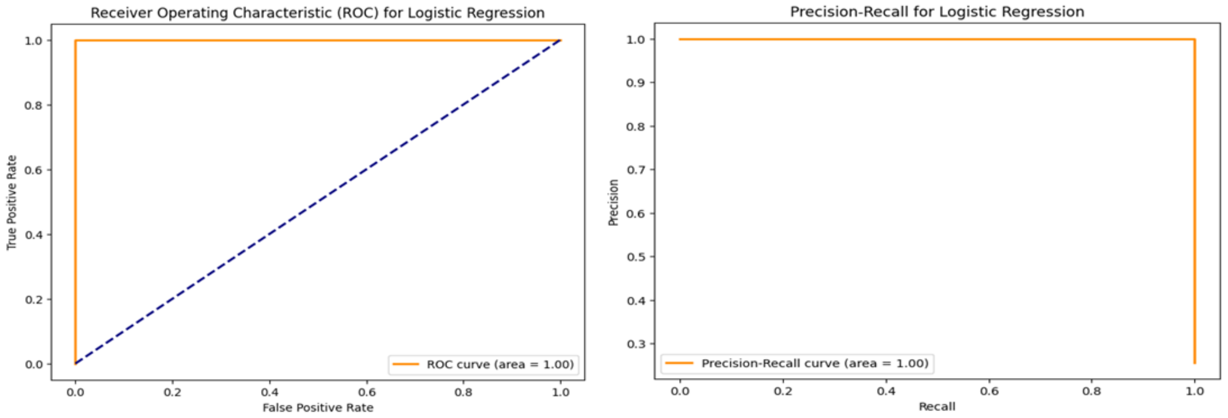


Figure 7. The x-axis Represents the False Positive Rate (FPR), which Measures the Proportion of Negative Instances Incorrectly Classified as Positive. In contrast, the y-axis represents the True Positive Rate (TPR), indicating the model's accuracy in identifying positive instances. The orange line on the plot reaches the top left corner, which suggests the model perfectly distinguishes between positive and negative classes. The AUC, displayed as 1.00, confirms this perfect classification. An AUC of 1.00 represents the ideal scenario where the model correctly classifies all instances without false positives or negatives. This ROC curve indicates that the logistic regression model is highly accurate. However, in real-world applications, such a perfect curve is rare and could indicate exceptional performance or possible overfitting, depending on the data.

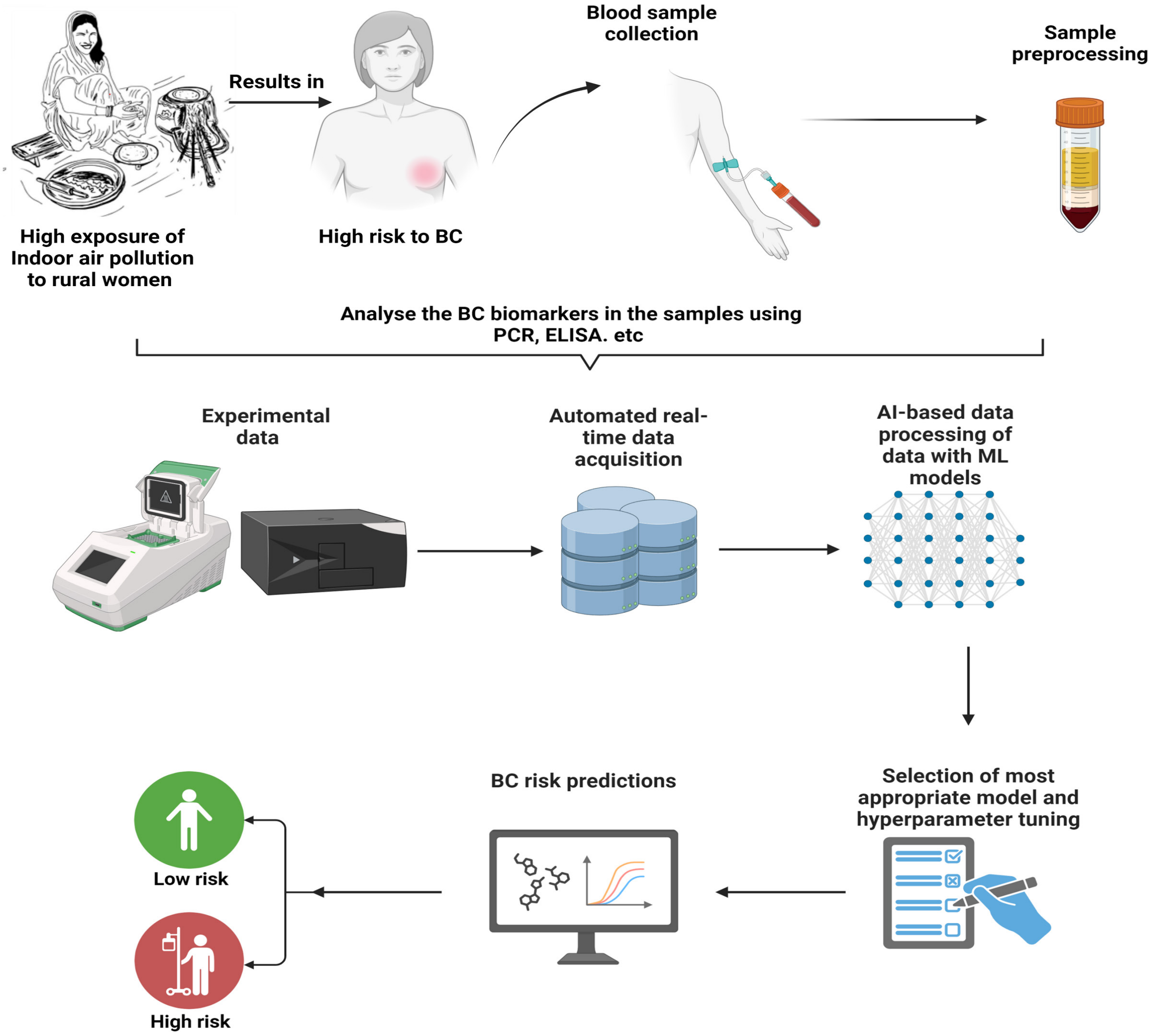


Figure 8. The Figure Illustrates the Comprehensive Impact of Our Study, commencing with blood collection, followed by processing, and culminating in the application of artificial intelligence to predict the risk of breast cancer.

levels, tumor growth, and hormonal receptor expression, warranting further investigation.

This study found a significant association between mtDNA methylation and TR expression, highlighting its potential role in BC risk, particularly with aging. Higher mtDNA methylation was linked to advanced age and increased BC risk. Additionally, exposure to PM2.5 was associated with abnormal DNA methylation patterns in healthy individuals, and early-life exposure to IAP correlated with methylation patterns in breast tumors [30]. There is a relationship between TR activity and mtDNA, suggesting physiological connections. Research indicates that telomere length (TL) and mtDNA are co-regulated, both related to BC risk[31]. Oxidative stress damages telomeric DNA, and antioxidants may help reduce this damage [32]. Our analysis showed a modest increase in AFP levels associated with mtDNA methylation, suggesting a link between mitochondrial DNA changes and BC. Increased levels of AFP may indicate possible mitochondrial dysfunction associated with cancer, and recent studies propose that mtDNA methylation could be a marker of overall cellular stress [33]. It is important to recognize that DNA methylation exhibits strong tissue specificity [34], influenced by cell-type composition, chromatin context, and local transcriptional activity. Consequently, methylation alterations detected in peripheral blood leukocytes may not accurately mirror those occurring within breast epithelial tissue, particularly in the context of tumor heterogeneity and distinct epigenetic regulation. While peripheral blood provides a practical and minimally invasive source for population-level screening, its use as a surrogate for breast tissue epigenetic profiling has limitations. Therefore, future research should include matched tissue-based analyses and functional validation and resolve cell-type specific methylation dynamics to confirm the biological relevance of peripheral methylation signatures in breast cancer.

Machine learning models are employed for BC classification to predict tumor malignancy based on features like DNA methylation, biomarkers from biopsies, imaging, and genetic tests. Various models, including Logistic Regression, Random Forest, Gradient Boosting, SVM, and KNN, are used to capture patterns in labeled data. Cross-validation ensures model reliability by splitting data into training and testing sets, and performance is assessed through metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Logistic Regression is is especially appreciated for its interpretability, aiding clinicians understand the impact of biomarkers on risk. Despite the model's impressive results, certain limitations must be acknowledged. The cross-sectional design restricts the ability to draw causal conclusions, and the small, age-discrepant group decreases statistical power and generalizability. Uncontrolled confounding factors such as hormonal status, menopause, genetics, and lifestyle may have impacted the outcomes. Although significant alterations in mitoepigenetics and non-coding RNA expression were noted, functional studies are necessary for mechanical validation. Furthermore, the machine learning model was created using a single dataset without external validation, which limits its clinical applicability. Future research should involve larger, longitudinal cohorts, incorporate multi-omics data, and implement robust model validation to enhance and broaden these findings. Additionally, the model's performance depends on the quality of the input data; thus, ongoing data updates and monitoring are essential to ensure sustained accuracy over time.

In conclusion, this research highlights a crucial connection between prolonged exposure to IAP, age-related modifications in mitochondrial epigenetics, and an increased likelihood of BC among rural women in India. Our findings show significant relationships between exposure to particulate matter, changes in mitochondrial DNA methylation, levels of inflammatory cytokines, and irregular expression of non-coding RNA, especially in older female participants. Multivariate Cox regression identified mtDNA methylation as a key biomarker for age-related BC risk. In addition, integrating molecular biomarkers with AI-optimized models demonstrates a high degree of predictive precision for BC risk, indicating promising potential for non-invasive early detection strategies that support personalized medicine by elucidating environmental cancer links, and guiding future preventive research (Figure 8). However, the cross-sectional design, limited sample size, and absence of tissue-specific validation calls for larger, longitudinal studies. Future work should include breast tissue epigenetic profiling, toxicological PM2.5 analysis, and multi-omics approaches to refine our understanding of environmental contributions to BC and advance personalized prevention strategies.

Author Contribution Statement

PKM devised the concept, developed the methodology, supervised the experiments and extramural project; PK, NS, and NN performed most experiments; NN and PR collected and characterized samples; VG developed the AI; NN, AC, and AKR prepared the tabulations; NS, NN and VG designed the figures; RT, AB, RKS, and PKM performed the data analysis and interpretation; and PK, VG, AC and PKM drafted the original manuscript.

Acknowledgements

The authors thank the technical staff members, Mr. Vinay Singh Raghwanshi and Mr. Sagar Patel, for their assistance in collecting samples.

Funding Statement

The authors would like to express their gratitude to the Indian Council of Medical Research (ICMR), the Department of Health Research (DHR), and the Ministry of Health and Family Welfare (MoHFW) of the Government of India in New Delhi for their financial support.

If any scientific body approved it/ if it is part of an approved student thesis

Indian Council of Medical Research (ICMR), Department of Health Research (DHR), Ministry of Health and Family Welfare (MoHFW), Government of India, New Delhi (IN).

Ethical Declaration

The study received approval from the Institutional Ethics Committee (IEC) of ICMR-NIREH under the reference number ITR-GIA/BC/2021-22.

Data Availability

The data supporting this study's findings are available upon request from the corresponding author.

Conflict of Interest

None.

Abbreviations

AFP: Alpha-fetoprotein
AG: Age
AI: Artificial Intelligence
AUC: Area under curve
BC: Breast Cancer
BER: Base excision repair
CEA: Carcinoembryonic Antigen
CM-H2DCFDA: 5-(and-6)-Chloromethyl-20, 70-dichlorodihydrofluorescein Diacetate Acetyl Ester
CO: Carbon Monoxide
ELISA: Enzyme Linked Immunosorbent Assay
FPG: Formamidopyrimidine Glycosylase
FPR: False Positive Rate
G-CSF: Granulocyte-Colony Stimulating Factor
HAP: Household Air Pollution
IAP: Indoor Air Pollution
IL: Interleukin
IL-6: Interleukin-6
ISR: Integrated Stress Response
KNN: K-Nearest Neighbors
mtDNA: Mitochondrial DNA
ncRNA: Non-coding Ribonucleic Acids
NF-κB: Nuclear Factor kappa B
PM: Particulate Matter
PM2·5: Particulate Matter with a diameter of 2.5 micrometers or smaller
ROC: Receiver Operating Characteristic

SEER: Surveillance, Epidemiology, and End Results
SVM: Support Vector Machine
TE: Total Exposure
TERT: Telomerase Reverse Transcriptase
TL: Telomere length
TFAM: Mitochondrial Transcription Factor A
TNF-α: Tumor Necrosis Factor Alpha
TPR: True Positive Rate

References

1. Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025;31(4):1154-62. https://doi.org/10.1038/s41591-025-03502-3.
2. Chowdhury S, Pillarisetti A, Oberholzer A, Jetter J, Mitchell J, Cappuccilli E, et al. A global review of the state of the evidence of household air pollution's contribution to ambient fine particulate matter and their related health impacts. Environ Int. 2023;173:107835. https://doi.org/10.1016/j.envint.2023.107835.
3. White AJ, Teitelbaum SL, Stellman SD, Beyea J, Steck SE, Mordukhovich I, et al. Indoor air pollution exposure from use of indoor stoves and fireplaces in association with breast cancer: A case-control study. Environ Health. 2014;13:108. https://doi.org/10.1186/1476-069x-13-108.
4. World health organization: Who. Household air pollution. 2024. Available from: Https://www.Who.Int/news-room/fact-sheets/detail/household-air-pollution-and-health.
5. Gabet S, Lemarchand C, Guénel P, Slama R. Breast cancer risk in association with atmospheric pollution exposure: A meta-analysis of effect estimates followed by a health impact assessment. Environ Health Perspect. 2021;129(5):57012. https://doi.org/10.1289/ehp8419.
6. Ravindra K. Emission of black carbon from rural households kitchens and assessment of lifetime excess cancer risk in villages of north india. Environ Int. 2019;122:201-12. https://doi.org/10.1016/j.envint.2018.11.008.
7. Niehoff NM, Terry MB, Bookwalter DB, Kaufman JD, O'Brien KM, Sandler DP, et al. Air pollution and breast cancer: An examination of modification by underlying familial breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2022;31(2):422-9. https://doi.org/10.1158/1055-9965.Epi-21-1140.
8. Balakrishnan K, Steenland K, Clasen T, Chang H, Johnson M, Pillarisetti A, et al. Exposure-response relationships for personal exposure to fine particulate matter (pm(2·5)), carbon monoxide, and black carbon and birthweight: An observational analysis of the multicountry household air pollution intervention network (hapin) trial. Lancet Planet Health. 2023;7(5):e387-e96. https://doi.org/10.1016/s2542-5196(23)00052-9.
9. Bhargava A, Shukla A, Bunkar N, Shandilya R, Lodhi L, Kumari R, et al. Exposure to ultrafine particulate matter induces nf-κβ mediated epigenetic modifications. Environ Pollut. 2019;252(Pt A):39-50. https://doi.org/10.1016/j.envpol.2019.05.065.
10. Bhargava A, Tamrakar S, Aglawe A, Lad H, Srivastava R, Mishra D, et al. Ultrafine particulate matter impairs mitochondrial redox homeostasis and activates phosphatidylinositol 3-kinase mediated DNA damage responses in lymphocytes. Environmental pollution (Barking, Essex : 1987). 2018;234:406-19. https://doi.org/10.1016/j.envpol.2017.11.093.
11. Abdel-Razeq H, Abu Rous F, Abuhijla F, Abdel-Razeq N, Edaily S. Breast cancer in geriatric patients: Current landscape and future prospects. Clin Interv Aging. 2022;17:1445-60. https://doi.org/10.2147/cia.S365497.
12. Agrawal D, Khetan R, Ratre P, Rehman A, Srivastava R, Reszka E, et al. Cell-free circulating mirnas-lncrnas-mrnas as predictive markers for breast cancer risk assessment in women exposed to indoor air pollution. Case Stud Chem Environ Eng. 2022;6:100267. https://doi.org/10.1016/j.cscee.2022.100267.
13. Bhargava A, Kumari R, Khare S, Shandilya R, Gupta PK, Tiwari R, et al. Mapping the mitochondrial regulation of epigenetic modifications in association with carcinogenic and noncarcinogenic polycyclic aromatic hydrocarbon exposure. Int J Toxicol. 2020;39(5):465-76. https://doi.org/10.1177/1091581820932875.
14. Mishra P, Bhargava A, Khetan R, Bunkar N, Chauhan P, Mukherjee S, et al. Integrated mitoepigenetic signalling mechanisms associated with airborne particulate matter exposure: A cross-sectional pilot study. Atmospheric Pollution Research. 2022;13:101399. https://doi.org/10.1016/j.apr.2022.101399.
15. Bunkar N, Sharma J, Chouksey A, Kumari R, Gupta PK, Tiwari R, et al. Clostridium perfringens phospholipase c impairs innate immune response by inducing integrated stress response and mitochondrial-induced epigenetic modifications. Cell Signal. 2020;75:109776. https://doi.org/10.1016/j.cellsig.2020.109776.
16. Shandilya R, Sobolev AM, Bunkar N, Bhargava A, Goryacheva IY, Mishra PK. Quantum dot nanoconjugates for immuno-detection of circulating cell-free mirnas. Talanta. 2020;208:120486. https://doi.org/10.1016/j.talanta.2019.120486.
17. Mishra P, Bunkar N, Singh R, Kumar R, Gupta P, Tiwari R, et al. Comparative profiling of epigenetic modifications among individuals living in different high and low air pollution zones: A pilot study from india. Environmental Advances. 2021;4:100052. https://doi.org/10.1016/j.envadv.2021.100052.
18. Shandilya R, Bhargava A, Ratre P, Kumari R, Tiwari R, Chauhan P, et al. Graphene quantum-dot-based nanophotonic approach for targeted detection of long noncoding rnas in circulation. ACS Omega. 2022;7(30):26601-9. https://doi.org/10.1021/acsomega.2c02802.
19. Mishra PK, Kumari R, Bhargava A, Bunkar N, Chauhan P, Tiwari R, et al. Prenatal exposure to environmental pro-oxidants induces mitochondria-mediated epigenetic changes: A cross-sectional pilot study. Environ Sci Pollut Res Int. 2022;29(49):74133-49. https://doi.org/10.1007/s11356-022-21059-3.
20. Zins K, Peka E, Miedl H, Ecker S, Abraham D, Schreiber M. Association of the telomerase reverse transcriptase rs10069690 polymorphism with the risk, age at onset and prognosis of triple negative breast cancer. Int J Mol Sci. 2023;24(3). https://doi.org/10.3390/ijms24031825.
21. Soni N, Kaur P, Gurjar V, Bhargava A, Tiwari R, Chouksey A, et al. Unveiling the interconnected dynamics of mitochondrial dysfunction associated with age-related cardiovascular risk: A cross-sectional pilot study. Cureus. 2025;17(4):e82961. https://doi.org/10.7759/cureus.82961.
22. Gurjar V, Nazeer N, Bhargava A, Soni N, Kaur P, Tiwari R, et al. Deep learning-enabled nanophotonic test leveraging poly-l lysine-tethered carbon quantum dots to assess the risk of cardiovascular disease. Microchem J. 2024;207:112164. https://doi.org/10.1016/j.microc.2024.112164.
23. Nazeer N, Gurjar V, Ratre P, Dewangan R, Zaidi K, Tiwari R, et al. Cardiovascular disease risk assessment through sensing the circulating microbiome with perovskite quantum

dots leveraging deep learning models for bacterial species selection. Mikrochim Acta. 2024;191(5):255. https://doi.org/10.1007/s00604-024-06343-y.

24. Kang H, Dai Z, Ma X, Ma L, Jin Y, Liu X, et al. A genetic variant in the promoter of ape1 gene (-656 t>g) is associated with breast cancer risk and progression in a chinese population. Gene. 2013;531(1):97-100. https://doi.org/10.1016/j.gene.2013.08.052.
25. Kumar H, Kumar RM, Bhattacharjee D, Somanna P, Jain V. Role of nrf2 signaling cascade in breast cancer: Strategies and treatment. Front Pharmacol. 2022;13:720076. https://doi.org/10.3389/fphar.2022.720076.
26. Hou L, Barupal J, Zhang W, Zheng Y, Liu L, Zhang X, et al. Particulate air pollution exposure and expression of viral and human micrornas in blood: The beijing truck driver air pollution study. Environ Health Perspect. 2016;124(3):344-50. https://doi.org/10.1289/ehp.1408519.
27. Bhattacharjee R, Prabhakar N, Kumar L, Bhattacharjee A, Kar S, Malik S, et al. Crosstalk between long noncoding rna and microrna in cancer. Cell Oncol (Dordr). 2023;46(4):885-908. https://doi.org/10.1007/s13402-023-00806-9.
28. Rea IM, Gibson DS, McGilligan V, McNerlan SE, Alexander HD, Ross OA. Age and age-related diseases: Role of inflammation triggers and cytokines. Front Immunol. 2018;9:586. https://doi.org/10.3389/fimmu.2018.00586.
29. Ma Y, Ren Y, Dai ZJ, Wu CJ, Ji YH, Xu J. Il-6, il-8 and tnf-α levels correlate with disease stage in breast cancer patients. Adv Clin Exp Med. 2017;26(3):421-6. https://doi.org/10.17219/acem/62120.
30. Callahan CL, Bonner MR, Nie J, Han D, Wang Y, Tao MH, et al. Lifetime exposure to ambient air pollution and methylation of tumor suppressor genes in breast tumors. Environ Res. 2018;161:418-24. https://doi.org/10.1016/j.envres.2017.11.040.
31. Robinson NJ, Schiemann WP. Telomerase in cancer: Function, regulation, and clinical translation. Cancers (Basel). 2022;14(3). https://doi.org/10.3390/cancers14030808.
32. Shen J, Gammon MD, Terry MB, Wang Q, Bradshaw P, Teitelbaum SL, et al. Telomere length, oxidative damage, antioxidants and breast cancer risk. Int J Cancer. 2009;124(7):1637-43. https://doi.org/10.1002/ijc.24105.
33. Villacampa MJ, Moro R, Naval J, Failly-Crepin C, Lampreave F, Uriel J. Alpha-fetoprotein receptors in a human breast cancer cell line. Biochem Biophys Res Commun. 1984;122(3):1322-7. https://doi.org/10.1016/0006-291x(84)91236-1.
34. Lowe R, Slodkowicz G, Goldman N, Rakyan VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics. 2015;10(4):274-81. https://doi.org/10.1080/15592294.2014.1003744.