AI-ENABLED RESPIRATORY RISK FORECASTING: LEVERAGING PUBLIC IAQ DATA AND BIOAEROSOL SIGNALS FOR PREDICTIVE HEALTH INTELLIGENCE

Cameron Jones[1,2,3] and Bhavya Raj[1]

[1] Biological Health Services, Australia

[2] National Institute of Integrative Medicine, Australia

[3] BioMedix Co. Ltd., Thailand

Abstract

Advances in AI, no-code development, and access to environmental data APIs are enabling scalable, customizable tools for forecasting respiratory health risk. This project describes an API-driven, web-based system that integrates real-time outdoor bioaerosol proxies including PM2.5, PM10, pollen, and other meteorological variables to model local illness risk without deploying new physical sensors. Grounded in research identifying inverse associations between bioaerosol concentrations and influenza-like illnesses (ILIs), our approach uses multiple public APIs and infodemiology techniques to deliver risk stratification via a green/yellow/red exposure model. Development followed a conversational AI-driven programming method called vibe coding, where dashboards were built using Replit, Lovable, Manus, and Perplexity Labs to handle front-end/back-end functionality. Large language models like ChatGPT, Claude, Gemini, DeepSeek, Grok and Perplexity were used to brainstorm logic, select frameworks, troubleshoot code, and validate design decisions. This iterative, multi-LLM workflow accelerated prototyping while enhancing robustness and flexibility. This work demonstrates how open environmental data, Google Trends surveillance, and no-code platforms can deliver accessible tools for respiratory risk communication. This is particularly valuable in regions like Australia and Southeast Asia where real-time mould monitoring is absent or not available publicly as an API. Grounded in the hygiene hypothesis, which suggests that exposure to non-pathogenic microbiota (e.g. mould spores, pollen) can condition immune responses, we model PM2.5, PM10, and pollen concentrations as real-time proxies for airborne microbial exposure. Early results for some cities like Melbourne show an inverse relationship between bioaerosol levels and Google Trends search interest in respiratory symptoms, supporting the use of these air quality signals for scalable, illness forecasting intelligence. Researchers, educators, and risk analysts can create dashboards for display, analysis, and forecasting.

Keywords: bioaerosols, vibe coding, infodemiology, mould spores

1. Introduction

Airborne bioaerosols including pollen and mould spores as well as particulate matter (PM2.5, PM10) play complex roles in respiratory health. Our foundational hypothesis was directly informed by studies showing that airborne mould spores and pollen are inversely correlated with influenza-like illness and COVID-19 presentations (Shah et al. 2021a, 2021b; Hoogeveen et al. 2021). Hoogeveen et al. (2021) demonstrated that in the Netherlands, pollen concentrations inversely predict flu-like illness, including COVID-19, with distinct thresholds marking epidemic transitions, while solar radiation acts as a co-inhibitor. Extending these observations, the Shah et al. group showed in Chicago that both pollen and mould spore counts inversely correlate with emergency department presentations for ILI and COVID-19, proposing that these bioaerosols compete with respiratory viruses for innate immune receptors such as Toll-like receptor 4 (TLR4).

Ladau et al. (2021) offered complementary evidence from mycobiome analyses across the United States, finding that greater indoor–outdoor fungal beta-diversity was associated with lower COVID-19 infection fatality ratios, suggesting environmental fungal exposures may condition immune responses. Climate change is projected to worsen respiratory risks via longer pollen seasons, more intense wildfire events, and altered precipitation patterns affecting mould growth (Gu et al. 2024).

While most research was conducted in the Northern Hemisphere, we hypothesised that similar seasonal bioaerosol effects may operate across Australia and Southeast Asia. However, an immediate challenge arises: unlike pollen monitoring, mould spore counts are rarely digitised or real-time. In Australia and Thailand, mould classification remains largely manual, conducted under microscopy by trained aerobiologists (Jones & Neumeister-Kemp). Consistent and accurate high-resolution mould data is largely absent from public health datasets. To address this limitation, we developed digital proxies for mould exposure using accessible real-time environmental indicators. PM2.5 and PM10 concentrations, combined with pollen counts and meteorological variables (humidity, temperature, dew point), were chosen as surrogate inputs. These variables co-vary with mould spore concentrations under certain climatic conditions and form a composite picture of the aerobiological environment affecting respiratory health.

The characterization of fungal contributions to airborne particulate matter such as PM2.5 and PM10 is significant for understanding air quality and its effects on human health. Studies have shown that fungi comprise a notable fraction of particulate matter in urban environments. It has been reported that fungal spores can contribute approximately 4% to 11% of the total mass in PM2.5 and account for about 60% of the coarse organic carbon in PM10 (Burshtein et al., 2011). Additionally, the contribution of fungal spores to PM10 mass has been estimated to range from 1.6% to as high as 18.2%, with a mean value reported at about 7.9% (Zhang et al., 2010).

In order to rapidly develop user interface dashboards we used vibe coding. This is an emergent paradigm in AI-assisted software development that leverages conversational interactions with large language models (LLMs) to facilitate intuitive and iterative code generation, as explored in recent studies (Sarkar & Drosses, 2025; Sapkota et al., 2025). In the context of our project, vibe coding was employed to rapidly prototype a respiratory risk visualisation system, enabling researchers with limited programming expertise to translate complex bioaerosol and infodemiology concepts into functional dashboards. Unlike traditional coding, which demands precise syntax, vibe coding abstracts low-level implementation details, allowing developers to focus on high-level intent and system design through natural language prompts (Sapkota et al., 2025). This approach aligns with Sarkar and Drosses (2025), who highlight vibe coding's iterative workflow, where developers engage in prompt-based dialogues with AI to refine code, integrating tools like Ambee, OpenWeatherMap APIs for real-time environmental data. By fostering a co-creative process, vibe coding democratizes development, making it accessible for non-programmers to build responsive, data-driven health communication tools.

2. Conceptual Framework

Our approach similarly, integrates environmental bioaerosol proxies with infodemiology in electronic media to inform public health policy (Eysenbach 2002). Porcu et al. (2023) demonstrated that Google Trends data for terms like "fever" and "cough" could reliably precede COVID-19 epidemic waves in Italy with >80% sensitivity. Similarly, Shih et al. (2024) applied this framework to influenza-like illness surveillance in Taiwan, showing that search queries significantly predicted ILI emergency visits when combined with meteorological predictors. Recent work by Chu et al. (2025) further reinforced the predictive power of Google Trends by demonstrating that merged search volumes for long COVID symptoms significantly enhanced the forecasting of long COVID prevalence, highlighting the broader applicability of infodemiological methods in real-time disease surveillance and risk communication. We extended this philosophy to individual-level risk communication by integrating environmental APIs, Google Trends proxies for ILI activity (including mould proxies), and local weather data to deliver accessible, location-specific health intelligence. The core objective was to develop a modular, API-based forecasting system that could quantify bioaerosol risk by integrating multiple public data sources. Recognizing that no dedicated mould spore APIs exist country-wide, we conceptualized using PM2.5/PM10 and pollen data as proxies for mould bioaerosol risk, supported by published research on their synergistic and inverse relationships with respiratory illness. This "bioaerosol risk" model frames PM and pollen not just as pollutants but as immunomodulatory environmental variables whose interactions can influence illness risk.

3. Bioaerosol Classifications Systems

Initial models utilized raw PM2.5 and PM10 values retrieved directly from API feeds, enabling real-time data integration without modifications. This approach prioritized stability and confidence in the data pipeline, ensuring reliable baseline measurements. Once the API integration was validated, we implemented advanced parameterizations to refine the analysis. For example, bioaerosol-specific coefficients were introduced to isolate biological components from total PM mass, enhancing the tool's precision for health-related insights. This iterative process - from raw PM data to specialized derivations exemplifies the dashboard's adaptability, allowing seamless transitions between broad environmental monitoring and targeted biological particle assessment.

The bioaerosol coefficients (0.024 for PM2.5 and 0.048 for PM10) used in some of our dashboards are derived from peer-reviewed studies measuring bioaerosol contributions to particulate matter (Liu et al., 2019; Humbert et al., 2011). These values represent the mass fractions of biological particles (e.g., bacteria, fungi, pollen) within PM2.5 (2.4 ± 1.9%) and PM10 (4.8 ± 3.2%), converted to decimal form. The foundational research employed direct fluorescent staining and microscopic imaging to quantify bioaerosols in ambient air samples, assuming spherical particles with a unit density of 1 g/cm³. While global averages suggest higher bioaerosol contributions (~16% for both PM sizes), the conservative values from urban Chinese studies (Liu et al., 2019) provide a robust baseline for our models. The QUT Allergy Research Group's scientifically validated thresholds form the basis of Australia's authoritative pollen risk classification system, as implemented by the AusPollen Brisbane monitoring site. Specifically calibrated for Australian environmental conditions and seasonal patterns, this system categorizes grass pollen levels into four tiers: Low (<20 grains/m³, green), Moderate (20-49 grains/m³, orange), High (50-99 grains/m³, red), and Extreme (≥100 grains/m³, purple).

4. Methods

4.1. Vibe Coding & API Integration Summary

Development leveraged a no-code/low-code vibe coding approach, using generative AI (ChatGPT, Manus, Claude, Grok, Gemini, DeepSeek) for conversational system design; while no code was in Lovable, Replit, Bolt including API integration, and interface logic. Environmental and public health data were sourced via:

- Ambee API: PM2.5, PM10, pollen concentrations (tree, grass, weed), temperature, humidity, precipitation, wind speed, NO2, CO for real-time environmental data across Australia and globally.
- City of Melbourne Open Data: Local PM and weather data
- OpenWeatherMap: PM, humidity, temperature, precipitation and meteorological variables
- Google Trends: respiratory symptom searches as ILI proxies using search volume data for keywords like: fever, cough, sore throat, muscle ache, fatigue.

These APIs enable construction of daily risk profiles without deploying new local physical sensors, instead exploiting deployed sensor networks.

2.3 Development and Prototyping Phases

Development followed an iterative, AI-assisted workflow. ChatGPT, Gemini, Grok, DeepSeek, Perplexity and Claude were used for general conversations around development. Manual checks were vital to confirm that real data was being retrieved by the API and displayed since most front end platforms tend to auto display synthetic data.

5. Results

Multiple dashboards were created using these tools and AI-enhanced workflows. We had best success taking ideas, prompts and code from Claude, Manus and Perplexity and Perplexity Labs to create fully functional prototypes working on real data. The two main dashboards shown in the Appendix and built on respectively Lovable and Replit successfully demonstrate:

- Real-time integration of environmental APIs for Australian cities (PM2.5, PM10, pollen, weather)
- Computation of vibe zone risk classifications (green, yellow, red) reflecting bioaerosol-informed thresholds
- Storage and retrieval of environmental readings in structured databases
- Responsive, user-friendly UI design for public communication
- Integration of Google Trends data for ILI surveillance at state and national levels Integration of Google Trends data for ILI surveillance at state and national levels

5.1. The Hygiene Hypothesis

One of the first tasks was to obtain PM data. The City of Melbourne have API's for multiple sensor networks across the city. As proof of concept, we can query this data for local trends and then do the same for Google Trends. This allowed us to see general trends and to generate risk rating visualisations. Then the process could be automated, and all data sets aggregated into one of two dashboards (see Appendix 1 and 2). It became evident quickly that trends reported in the Victorian Department of Health 2025, Victorian respiratory surveillance reports regarding influenza-type illnesses including COVID were increasing. We could then explore these patterns for other cities and states. In Figure 1, we were able to quickly model Melbourne's open-source sensor network and average all the sensors to create a PM trend plot (Figure 1) showing the decline in PM over the period.

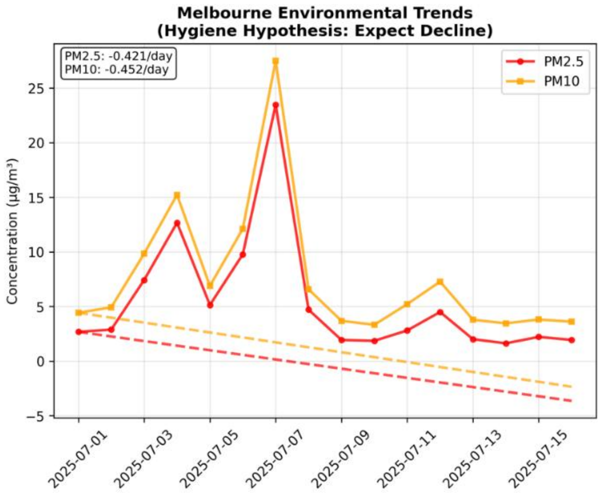


Figure 1. Validation in Manus.ai of a decline in PM2.5 and PM10 over a 4-week time frame (hygiene hypothesis) averaged across data from 9 sensors.

Different types of visualisations were possible which help with interpretation, data analysis and research (e.g. Figure 2).

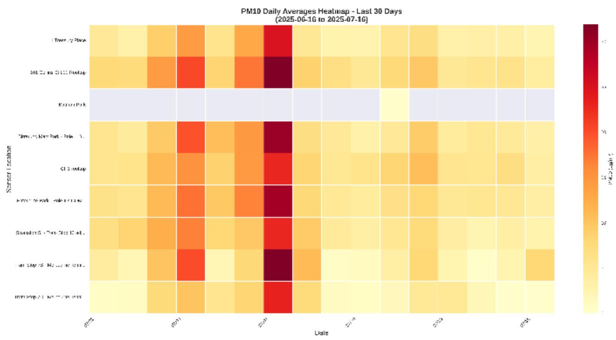


Figure 2. Heat map visualisation showing a decline in total particulate matter, PM10 correlation between Google trend data and discrete sensors, but despite sensor location variance, general trends showing a decline in PM is evident.

Then in Figure 3 a correlation matrix is shown that is helpful for proof of concept around digital mould proxies for PM and Google illness trend searches. Figure 4 shows a Perplexity Labs fragment displaying the time series of hourly Google Trends interest (0–100) for five health symptoms across Australia for a windowed period, July 10–17, 2025. The lines of best fit regression validate an increase in search volume for this windowed period. Each line in Figure 4 represents a symptom, showing rising search interest for fever (+73%), headache (+43%), and muscle pain (+25%), with cough stable and fatigue slightly decreasing. Statistically significant trends were observed for fever, headache, and muscle pain, demonstrating the utility of infodemiology for real-time health surveillance.

Problems we encountered during the process included: prompt fatigue, memory gaps and command misalignment. Model stability could be enhanced by regular session resets, prompt management and learning the role of code history and its relationship to GitHub and setting clear tasks with appropriate context.

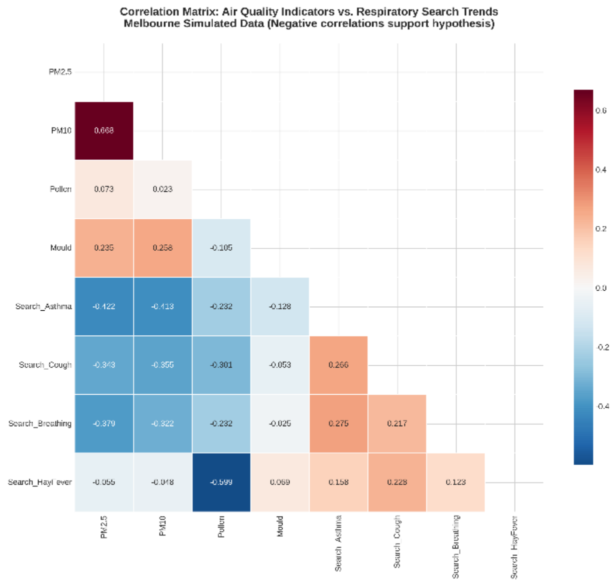


Figure 3. Air quality data correlation matrix.

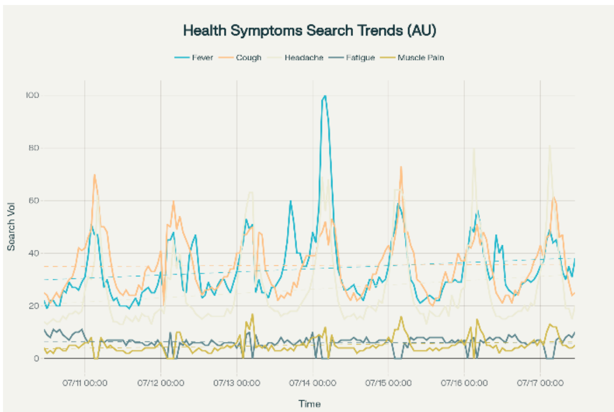


Figure 4. Perplexity Labs fragment of the hourly Google Trends interest (0–100) for five health symptoms in Australia over 7-days across Australia for a windowed period, July 10–17, 2025.

6. Discussion

This API-based approach offers an accessible, scalable alternative to traditional air quality monitoring systems that rely on costly, spatially sparse sensor networks. By leveraging publicly available environmental APIs and integrating bioaerosol-informed modelling, this system enables real-time respiratory risk visualisation without requiring new physical infrastructure. A central aim of this paper has been to explore and operationalize the observed inverse relationships between ambient bioaerosols - particularly pollen and fungal spores and influenza-like illness (ILI), as reported in foundational studies. While mould spores can be readily measured using spore traps, such data are typically collected in the context of indoor environmental assessments (e.g., water damage investigations) and remain largely siloed from public health applications. Moreover, mould spore concentrations are infrequently accessible via open APIs, limiting their integration into dynamic risk models. To address this, we sought to identify and test surrogate environmental indicators such as PM2.5, PM10, and pollen counts that correlate with bioaerosol load and may serve as viable proxies in the absence of direct mould spore measurements.

Importantly, we emphasise that we do not suggest high mould levels are protective. Rather, in the absence of accurate, location-specific mould data, these digital proxies for the outdoor air environment offer a viable epidemiological surrogate. The hygiene hypothesis, originally proposed to explain rising rates of allergies and autoimmune diseases in overly sanitized environments, posits that moderate exposure to diverse, non-pathogenic microbes such as those in outdoor bioaerosols like mould spores and pollen (Oh et al., 2024) helps train and modulate the immune system, potentially reducing susceptibility to certain infections by promoting balanced innate immune responses (e.g., via competition for receptors like TLR4). In the context of our work, this manifests in outdoor air dynamics: as bioaerosol proxies (e.g., PM2.5, PM10, and pollen counts) decline to low levels, we observe a corresponding uptick in influenza-like illness (ILI) indicators via Google Trends searches for symptoms like fever and cough, suggesting that ambient environmental exposures may create "protective zones" by inhibiting viral transmission through ecological competition or immune priming, as evidenced in foundational studies from the Netherlands and Chicago (Hoogeveen et al., 2021; Shah et al., 2021a, 2021b). This inverse relationship aligns with the hypothesis by framing background outdoor bioaerosols as beneficial surrogates for microbial diversity, where their absence signals vulnerability windows for respiratory viruses, particularly in seasonal patterns across Australian cities and states as tracked by our dashboards. However, this outdoor benefit contrasts sharply with indoor air scenarios, where elevated mould levels often stemming from water damage or poor ventilation correlate with worsened occupant symptoms, including respiratory irritation, asthma exacerbations, and allergic responses, as documented in allied investigations (Jones & Neumeister-Kemp, 2025). The apparent contradiction is reconciled by considering dose, context, and composition: outdoor exposures typically involve transient, low-concentration, diverse fungal communities that mimic natural ecological priming without overwhelming the system, whereas indoor mould amplifies risks through persistent, high-density growth of potentially toxigenic species, leading to chronic inflammation rather than protection; thus, our visualisation and forecasting tools emphasize outdoor proxies for population-level ILI risk stratification while advocating for indoor remediation to mitigate localized health hazards.

Our broader vision is to curate open, API-based environmental data streams for a unified respiratory health dashboard, providing immediate, historical, and forecast risk estimates tailored to geolocations. Ultimately, we aim to create a publicly accessible tool for proactive health decisions and population-level surveillance. This project underscores the value of no-code/low-code platforms and LLM-based code generation for rapid, collaborative development, enabling functional dashboards in hours to days and empowering non-programmers. However, API reliance is limited by geographic coverage and data quality. Google Trends infodemiology complements this, especially amid limited outdoor mould insights. Future work includes integrating local epidemiological data, deploying IoT sensors for hyperlocal indoor/outdoor sensing, and partnering with health authorities.

7. Conclusions

We present a flexible, API-driven, no-code-enabled respiratory risk visualisation system that integrates real-time air quality data or bioaerosol proxies and infodemiology surveillance to communicate dynamic, geo-located health risks. Vibe coding enabled rapid functional prototypes to be built quickly. This approach offers a scalable model for public health communication, requiring no new local sensor deployments while remaining extensible to future IoT integration. Our work demonstrates how environmental health analytics can be democratized through open data, AI-based code generation, and no-code platforms, supporting communities in anticipating and mitigating respiratory illness risks. The hygiene hypothesis may be extended to digital mould proxies, offering pragmatic, scalable toolsets for respiratory illness visualisation, personal monitoring and forecasting. This is considered particularly valuable for at-risk populations with asthma, COPD, or immunocompromise who face heightened risks from bioaerosol exposure, such as from mould-induced allergic reactions, infections, exacerbations, and chronic inflammatory responses.

Acknowledgments

We thank the data providers including Ambee and the City of Melbourne for environmental data access.

References

Shah R.B., Shah R.D., Retzinger D.G., Retzinger A.C., Retzinger D.A. & Retzinger G.S. 2021a, 'Confirmation of an inverse relationship between bioaerosol count and influenza-like illnesses, including COVID-19. On the contribution of mold spores', medRxiv preprint.

Shah R.B., Shah R.D., Retzinger D.G., Retzinger A.C., Retzinger D.A. & Retzinger G.S. 2021b, 'Competing bioaerosols may influence the seasonality of influenza-like illnesses, including COVID-19. The Chicago experience', Pathogens 10:1204.

Hoogeveen M.J., van Gorp E.C.M. & Hoogeveen E.K. 2021, 'Can pollen explain the seasonality of flu-like illnesses in the Netherlands?', Science of the Total Environment 755:143182.

Ladau J., Abuabara K., Walker A.M., Joachimiak M.P., Bansal I., Wu Y., Hoffman E.B., Kuo C., Falco N., Streich J., van der Laan M.J., Wainwright H.M., Brodie E.L., Hess M., Jacobson D. & Brown J.B. 2021, 'The impact of environmental mycobiomes on geographic variation in COVID-19 mortality', medRxiv preprint.

Gu W., Xie D., Li Q., Feng H., Xue Y., Chen Y., Tang J., Zhou Y., Wang D., Tong S. & Liu S. 2024, 'Association of humidity and precipitation with asthma: a systematic review and meta-analysis', Frontiers in Allergy 5:1483430.

Jones, C.L. & Neumeister-Kemp, H.G. (2025). The 1000 spores/m³ mould threshold revisited: perspectives and reinforcement from post-2023 literature. Journal of Bacteriology & Mycology: Open Access, 13(2):77–84.

Burshtein, N., Lang-Yona, N. and Rudich, Y., 2011. Ergosterol, arabitol and mannitol as tracers for biogenic aerosols in the eastern Mediterranean. *Atmospheric Chemistry and Physics*, 11(2):829–839.

Zhang, T., Engling, G., Chan, C.Y., Zhang, Y.N., Zhang, Z.S., Lin, M., Sang, X.F. and Li, Y.S., 2010. Contribution of fungal spores to particulate matter in a tropical rainforest. *Environmental Research Letters*, 5(2), p.024010.

Sarkar, A. & Drosos, I. (2025). Vibe coding: Programming through conversation with artificial intelligence. arXiv preprint.

Sapkota, R., Roumeliotis, K.I.& Karkee, M. (2025). Vibe coding vs. Agentic cding: Fundamenals andf ractical implications of agentic AI. arXiv preprint,

Eysenbach G. 2002, 'Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet', Journal of Medical Internet Research 4(1):e11.

Porcu G., Chen Y.X., Bonaugurio A.S., Villa S., Riva L., Messina V., Bagarella G., Maistrello M., Leoni O., Cereda D., Matone F., Gori A. & Corrao G. 2023, 'Web-based surveillance of respiratory infection outbreaks: retrospective analysis of Italian COVID-19 epidemic waves using Google Trends', Frontiers in Public Health 11:1141688.

Shih D.-H., Wu Y.-H., Wu T.-W., Chang S.-C. & Shih M.-H. 2024, 'Infodemiology of influenza-like illness: Utilizing Google Trends' big data for epidemic surveillance', Journal of Clinical Medicine 13(7):1946.

Chu, A.M.Y., Tsang, J.T.Y., Chan, S.S.C., Chan, L.S.H. and So, M.K.P., 2025. Utilizing Google Trends data to enhance forecasts and monitor long COVID prevalence. Communications Medicine, 5, Article 179.

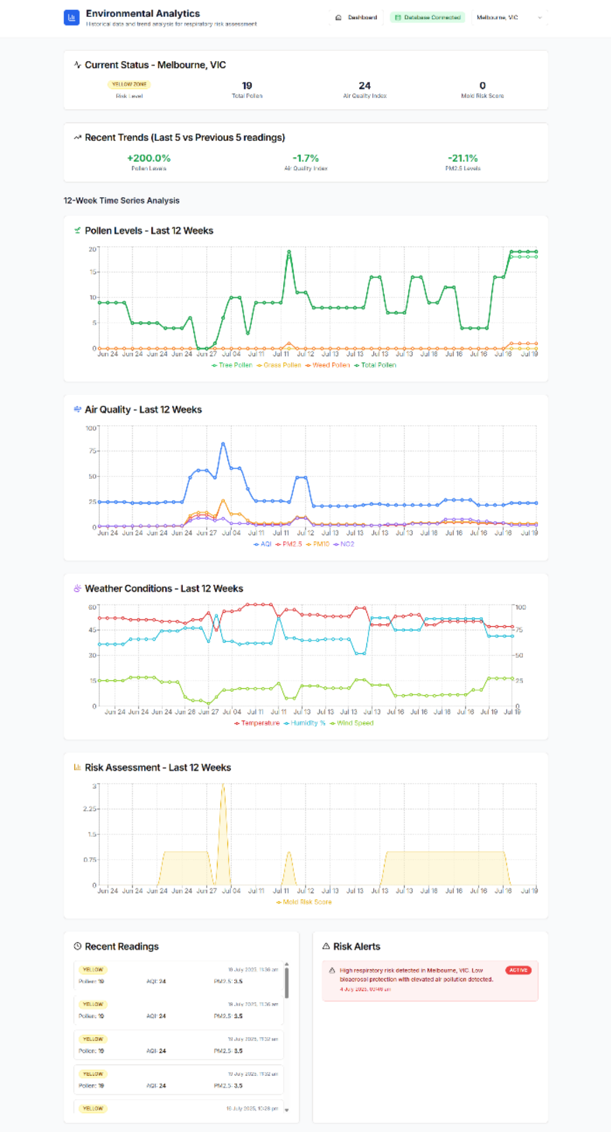
Liu, T. et al. (2019) ‘Bioaerosol concentrations and size distributions during the autumn and winter seasons in an industrial city of Central China’, Aerosol and Air Quality Research, 19(5), pp. 1095–1104.

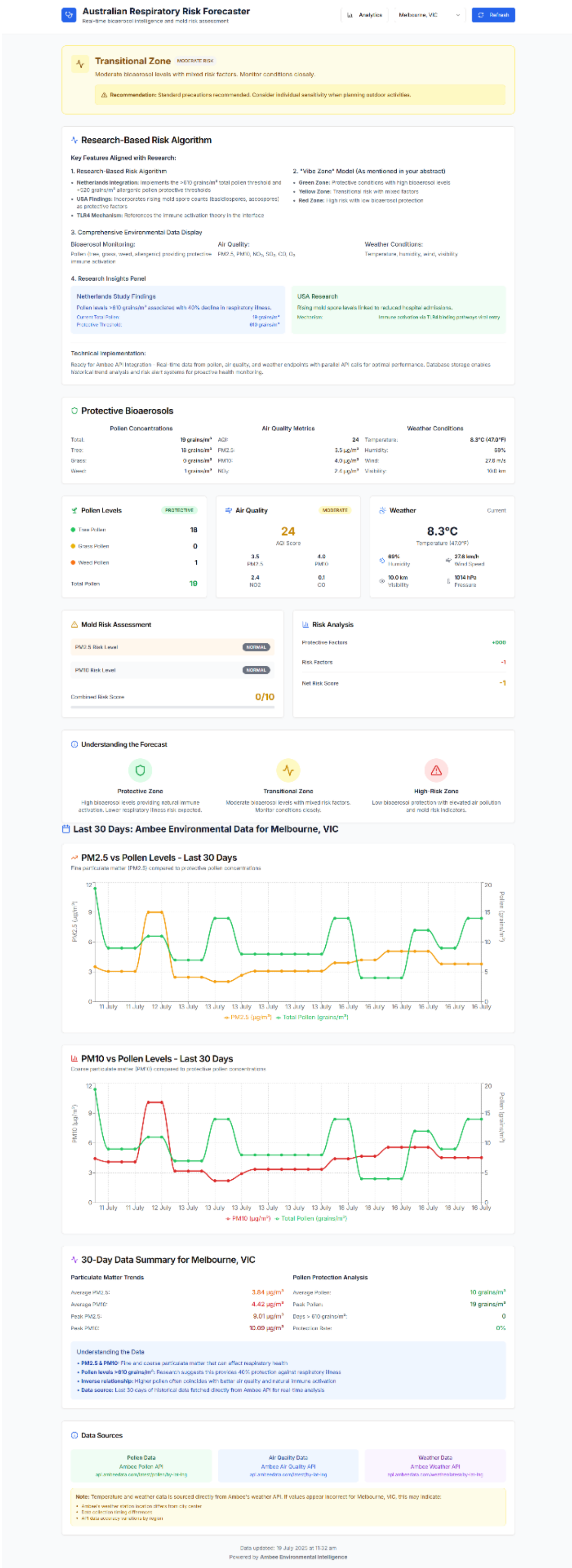
Humbert, S. et al. (2011) ‘Intake fraction for particulate matter: Recommendations for life cycle impact assessment’, Environmental Science & Technology, 45(11), pp. 4808–4816.

Oh J.W., Choi Y.J. & Lee K.S. 2024, 'Inverse trend between tree pollen and fungal concentrations with allergic sensitization rates in Seoul for 25 years', Allergy Asthma & Immunology Research 16(6):571–584.

Victorian Department of Health 2025, Victorian respiratory surveillance report: 11 July 2025, Victorian Government, Melbourne.

Appendix 1. Prototypes of the Australian Respiratory Risk Forecaster dashboard built via vibe coding in Replit. This system forecasts bioaerosol-informed respiratory risks using Ambee API data for PM2.5, PM10, pollen levels, air quality, and weather conditions in all 8 Capital cities of Australia, with risk assessments (e.g., "Protective Zone" classifications) and 30-day trend graphs comparing particulate matter to pollen concentrations. Drawing on Sarkar and Drosses' (2025) empirical findings, the development process involved conversational prompting workflows combining AI-generated code for real-time data fetching with manual overrides for context-specific adjustments highlighting vibe coding's role in material disengagement from low-level implementation while emphasizing expertise in rapid output verification and trust-building in AI-assisted health intelligence tools.





Appendix 2. Prototype dashboard UI of the Australian ILI Surveillance Dashboard, developed using vibe coding in Lovable. This interactive tool integrates Google Trends data for influenza-like illness (ILI) symptoms (e.g., fever, cough, headache, fatigue, muscle pain) across Australian states, displaying real-time indices, 4-week trends, and time-series visualizations. As analyzed by Sarkar and Drosses (2025), vibe coding enabled rapid prototyping through iterative goal-satisfaction cycles, where natural language prompts to LLMs (e.g., for API integration and UI refinement) alternated with quick code evaluation and hybrid debugging, redistributing expertise from manual syntax to high-level intent management and data visualization.