*The Laryngoscope*


# Computer-Aided Diagnosis of Maxillary Sinus Anomalies: Validation and Clinical Correlation

Debayan Bhattacharya, MSc; Benjamin Tobias Becker, MD; Finn Behrendt, MSc; Dirk Beyersdorff, MD, PhD; Elina Petersen, MSc; Marvin Petersen, MD; Bastian Cheng, MD, PhD; Dennis Eggert, PhD; Christian Betz, MD, PhD; Alexander Schlaefer, PhD; Anna Sophie Hoffmann, MD, PhD

**Objective:** Computer aided diagnostics (CAD) systems can automate the differentiation of maxillary sinus (MS) with and without opacification, simplifying the typically laborious process and aiding in clinical insight discovery within large cohorts.

**Methods:** This study uses Hamburg City Health Study (HCHS) a large, prospective, long-term, population-based cohort study of participants between 45 and 74 years of age. We develop a CAD system using an ensemble of 3D Convolutional Neural Network (CNN) to analyze cranial MRIs, distinguishing MS with opacifications (polyps, cysts, mucosal thickening) from MS without opacifications. The system is used to find correlations of participants with and without MS opacifications with clinical data (smoking, alcohol, BMI, asthma, bronchitis, sex, age, leukocyte count, C-reactive protein, allergies).

**Results:** The evaluation metrics of CAD system (Area Under Receiver Operator Characteristic: 0.95, sensitivity: 0.85, specificity: 0.90) demonstrated the effectiveness of our approach. MS with opacification group exhibited higher alcohol consumption, higher BMI, higher incidence of intrinsic asthma and extrinsic asthma. Male sex had higher prevalence of MS opacifications. Participants with MS opacifications had higher incidence of hay fever and house dust allergy but lower incidence of bee/wasp venom allergy.

**Conclusion:** The study demonstrates a 3D CNN's ability to distinguish MS with and without opacifications, improving automated diagnosis and aiding in correlating clinical data in population studies.

**Key Words:** Convolutional Neural Network, deep learning, maxillary sinus, Paranasal sinus, Population study.

**Level of Evidence:** 3

***Laryngoscope*, 134:3927–3934, 2024**

## INTRODUCTION

Morphological changes in paranasal sinuses, often seen in magnetic resonance imaging (MRI) scans and computed tomography (CT), have been extensively studied. Research shows a correlation between these changes and factors like allergies and smoking habits.[1] Furthermore, the relationship amongst patients with and without sinus opacifications have been explored in different patient groups: both symptomatic and asymptomatic,[2] exclusively symptomatic,[3] only asymptomatic,[4] and nonselected[5] individuals. However, in all the aforementioned studies, clinicians manually reviewed multiple MRI or CT slices to determine the presence or absence of paranasal opacification.

This manual process can strain clinicians and escalate their workload, leading to fatigue and potential misdiagnoses.[6] Deep learning (DL)-based computer-aided diagnosis system (CAD) presents an opportunity to enhance diagnostic accuracy and alleviate clinician workload by automating the classification of incidental findings. CNNs have demonstrated effectiveness in various aspects of paranasal opacification analysis, including screening, sinusitis classification, and tumor subtype differentiation. Existing studies often employ a two-stage methodology, first localizing sinuses and then classifying anomalies. For example, one study cropped x-ray images to classify anomalies[7] but failed to distinguish between left and right maxillary sinus anomalies. Another study segmented CT images and classified anomalies,[8] demanding pixel-level annotations for localization. Alternatively, a different approach used a CNN to detect key slices within CT images containing maxillary sinus volumes and subsequently classify maxillary sinus anomalies.[9] In our prior work, we explored unsupervised learning[10], contrastive learning[11], and multiple instance ensembling,[12] with multiple instance ensembling yielding the most promising results.



From the Institute of Medical Technology and Intelligent Systems (D.B., F.B., A.S.), Technische Universitaet Hamburg, Hamburg, Germany; Department of Otorhinolaryngology, Head and Neck Surgery and Oncology (D.B., B.T.B., D.E., C.B., A.S.H.), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Clinic and Polyclinic for Diagnostic and Interventional Radiology and Nuclear Medicine (D.B.), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Population Health Research Department, University Heart and Vascular Center (E.P.), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; and the Department of Neurology (M.P., B.C.), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Additional supporting information may be found in the online version of this article.

Editor's Note: This Manuscript was accepted for publication on March 13, 2024.

The authors have no funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Debayan Bhattacharya, Gebaude E, Raum 3.086, 3. Etage, Institute of Medical Technology and Intelligent Systems Technische Universitaet Hamburg. Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany. Email: debayan.bhattacharya@tuhh.de

**DOI: 10.1002/lary.31413**

We implemented an ensemble model (EM) of the multiple instance ensembling approach[12] based on the hypothesis that ensembling will improve model performance even more. We evaluated this EM on an unlabeled cohort within our population dataset to classify the presence or absence of opacification within the MS. Subsequently, we demonstrate a potential application for this EM which is rapid clinical insights of expanding cohorts which is typical in prospective population studies. To this end, we assessed the model's predictions by examining its relationships with various available factors, including smoking habits, alcohol consumption, BMI, chronic asthma, chronic bronchitis, sex, age, leukocyte count, highly sensitive C-reactive protein, and allergies and further assess if the discovered associations using our DL-based CAD are in accordance with previous literature which performed manual diagnosis. Overall, our research contributes to a better understanding of the prevalence of paranasal anomalies and further substantiates the potential of DL-based CAD to automate and enhance the efficiency of population studies, which have traditionally relied on labor-intensive methodologies.

## METHODS

### *Study Design and Participant Collective*

Hamburg City Health Study (HCHS)[13] is a single-center, prospective, population-based cohort study. A subsample ($N = 2619$) of the planned 45000 had cranial MRI scans recorded. MRI scans were recorded between February 08, 2016, and November 30, 2018, on individuals aged 45–74 years. Images were acquired using a 3-T Siemens Skyra MRI scanner (Siemens, Erlangen, Germany). 3D T2-weighted fluid attenuated inversion recovery (FLAIR) images were measured with the following sequence parameters: TR = 4700 ms, TE = 392 ms, 192 axial slices, ST = 0.9 mm, and IPR = 0.75 × 0.75 mm. Participant data included laboratory measurements of leukocytes/μL (LK) and high-sensitivity CRP (hCRP). Additionally, participants completed self-reported questionnaires documenting alcohol consumption per day (unit: grams/day [g/day]), smoking habits, diagnosis of chronic bronchitis or chronic obstructive pulmonary disease (COPD), and diagnosis of allergic bronchial asthma. Additionally, BMI, age, sex, and allergies of each participant were also recorded. The dataset comprised 2619 participants (56.05% men, 43.95% women) with a mean age of 63.98 (SD 8.32) years. Among these, 1069 participants (56.64% men, 43.36% women) with a mean age of 63.90 (SD 8.25) years were manually annotated to train a CNN. Among the annotated participants, 489 exhibited no opacifications in both left and right MS, while 580 showed at least one MS with polyp, cyst, or mucosal thickening (mucosa thickening >2 mm) opacification. These diagnoses were established by two ENT specialists and one ENT specialized radiologist. Figure 1 shows the flowchart of our study. Figure S1 shows exemplary MRIs exhibiting different pathologies.

### *DL Training, Validation and Test Dataset*

Detailed explanation of our data processing pipeline is reported in supplementary material sections 1, 2, and 3. We extract 30 MS volumes from each participant. Our dataset used to train the 3D CNN consisted of 19215 (59.91%) MS exhibiting no opacifications, 4815 (15.01%) MS exhibiting mucosal thickening, 6315 (19.69%) MS containing polyps in MS, 1185 (3.69%) MS exhibiting cyst opacification, and 540 (1.68%) MS containing polyps or cysts encompassing the entire MS volume. While constructing the test set, we considered two criteria: (i) accurate representation of opacifications in training, validation, and test set (ii) multiple volumes extracted from each participant does not occur in training, validation, or test set simultaneously. The test set contained 30% of the overall dataset. Table I contains the overall statistics of the training, validation, and test dataset.

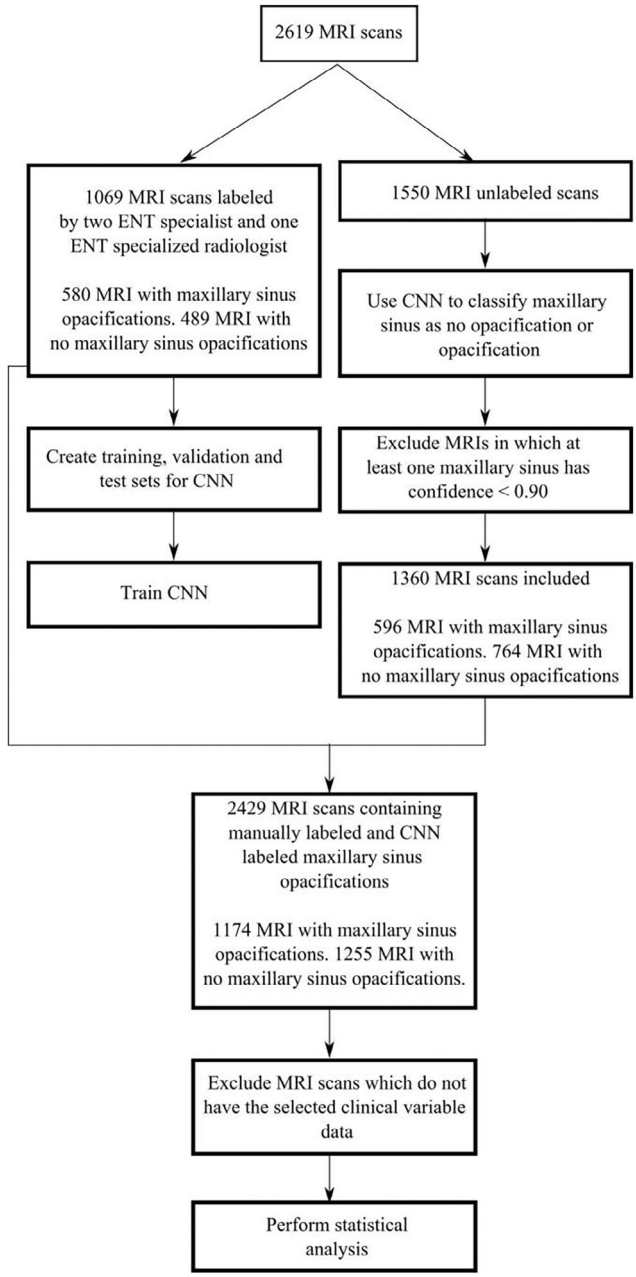


Fig. 1. Flowchart of our study.

### *Development of 3D Convolutional Neural Network*

Our CNN is a 3D implementation of a 264-layer convolutional neural network called DenseNet.[14,15] Figure 2 and Figure S2 shows illustrations of our deep learning method.

TABLE I.
Distribution of Opacifications in Dataset. The Percentages Are Reported Within Parenthesis.

| Dataset | Normal | Mucosal thickening | Polyp | Cysts | Fully occupied |
|---|---|---|---|---|---|
| Train | 10740 (59.86%) | 2700 (15.05%) | 3540 (19.73%) | 660 (3.67%) | 300 (1.66%) |
| Validation | 2700 (60%) | 675 (15%) | 885 (19.66%) | 165 (3.66%) | 75 (1.66%) |
| Test | 5775 (59.96%) | 1440 (14.95%) | 1890 (19.62%) | 360 (3.73%) | 165 (1.71%) |

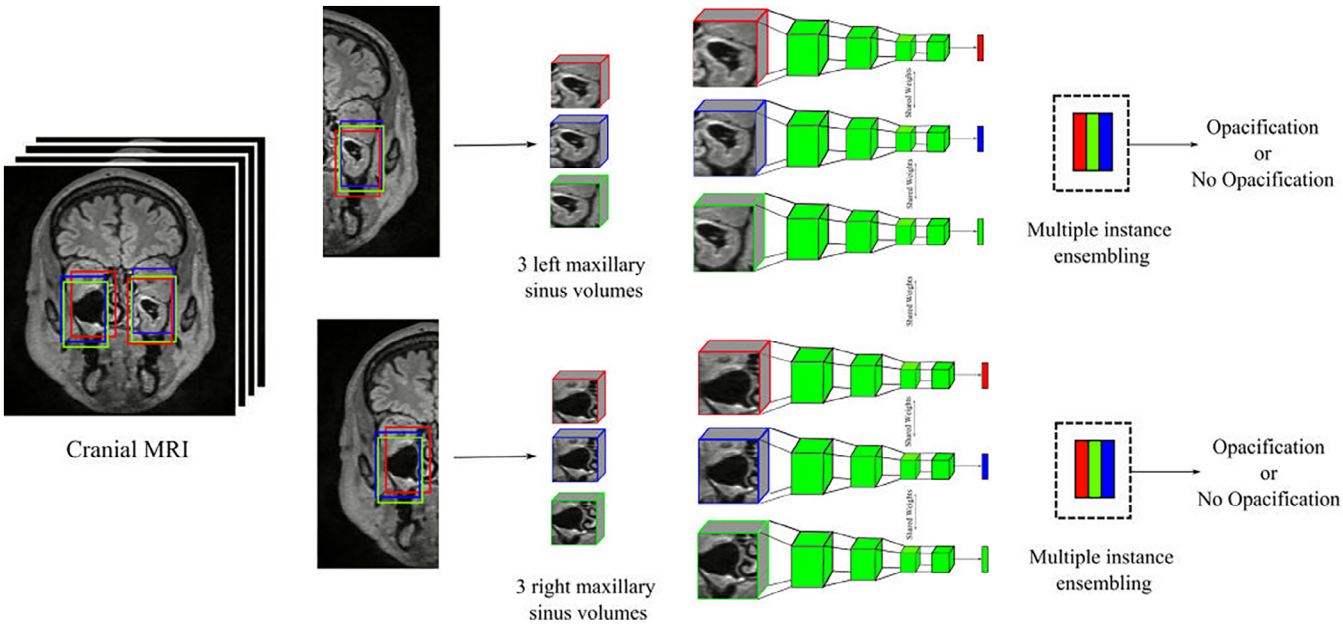


Fig. 2. Data processing pipeline showing how a single MRI is processed for inferring opacifications within the left and right MS of a single patient. As an illustration, 3 MS volumes are extracted from left and right side of MRI. Confidence score of 3 extracted MS volumes from the respective side are ensembled to create a single confidence score. In practice, we extract 15 MS volumes from left and right side of the MRI. The ensembling strategy is termed as multiple instance ensembling.

Implementation details are available in supplementary material section 3.1.

### *Training Protocol*

Our deep learning pipeline is discussed in the supplementary material section. The MS volumes are input into a 3D CNN for classification into two classes: "opacification" or "no opacification". The "opacification" class encompasses MS exhibiting mucosal thickening, polyps, or cysts. To ensure robustness, we employ a three-fold cross-validation strategy, training three distinct 3D CNNs. Subsequently, an EM is constructed by aggregating the predictions of these three 3D CNNs, and the final predictions are obtained by averaging the predictions of these models.

### *Statistical Analysis*

The McNemar test was used to test for statistically significant differences between the predictions of two 3D-CNN classifiers. We checked for statistical significance between 3D CNNs trained using single fold against EM. We performed $\chi^2$ test to check for statistically significant associations between categorical variables. We checked the point-biserial correlation, which is a special case of Pearson correlation, to measure the relationship between a continuous variable (BMI, alcohol consumption) and a dichotomous variable (participants with MS opacifications and participants without MS opacifications). A two-sided paired $p$-value <0.05 was considered significant. McNemar test was performed using MLxtend library[16] version 0.22.0, $\chi^2$ and point-biserial correlation was performed using SciPy[17] version 1.2.1. Python version 3.8.10 was used for our experiments.

## RESULTS

### *Performance Evaluation on Test Set*

The outcomes of our analysis are presented in Table II, where "CV" represents the cross-validation set, and all reported results pertain to the test set. The analysis reveals that ensembling contributes noticeably to the classification of MS anomalies, yielding an AUROC of 0.95, precision of 0.85, sensitivity of 0.85, and specificity of 0.90. In contrast, second-best model based on AUROC, a 3D-CNN trained on the first fold, achieves an AUROC of 0.93 in the same task. Using McNemar test, $p = 0.15$, $p = 0.01$, and $p = 6.8 \times e^{-4}$ was computed using the predicted labels of CV 1, CV 2, and CV 3 against the

predicted labels of EM, respectively. We achieve an accuracy of 0.90 for MS without opacification, 0.83 for MS with mucosal thickening, 0.88 for MS with polyp, 0.75 for MS with cyst, and 1.0 for fully occupied MS using EM on the test set.

### *Class Activation Maps*

Figure 3 presents activation maps exemplifying various scenarios of MS with and without associated opacifications. These activation maps generated using Guided Grad CAM[18] predominantly localize in clinically significant regions. Specifically, when the 3D-CNN classifier detects no anomalies, the activation maps are primarily concentrated along the boundaries of mucosal walls. In contrast, for cases of MS with opacifications, the activation maps tend to be focused within the mass of the opacifications.

### *Characteristics of Participants with Opacification and Without Opacification*

The objective of this experiment was to demonstrate a practical application of CNNs in automated diagnosis and the expedited generation of clinical insights through correlation studies between expanding cohorts in population studies, aiming to identify pertinent considerations. To simulate an expanding cohort, we leveraged our unlabeled dataset. The EM, trained on our labeled dataset ($N = 1069$), was applied to our unlabeled dataset ($N = 1550$). We implemented an inclusion criterion to select participants for further analysis by excluding MRIs where the EM confidence was <0.90 for at least one of the MS. From 1550 unlabeled MRI scans, only 1360 MRI scans satisfied the inclusion criteria. To both the groups, participants from the labeled dataset were added. Combining the labeled and unlabeled MRI scans, we considered 2429 MRIs for subsequent processing. A schematic drawing of our data handling strategy is presented in Figure 1. The 2429 MRIs were divided into two cohorts. The first cohort comprised participants for whom left and right MS had "no opacification" which is our *control* group. The second cohort consisted of participants with at least one MS having opacification which is our *case* group. Subsequently, we conducted a comparative analysis of these two groups, utilizing both self-evaluated questionnaires and blood reports available from the HCHS for each participant. Detailed results of these comparisons are presented in Tables III and IV. *case* group showed a higher mean alcohol consumption (20.65 (95% [CI] −34.83–76.15) vs. 15.56 (95% [CI] −30.18–61.31); $p < 0.001$) and BMI (27.11 (95% [CI] 18.99–35.33) vs. 26.26 (95% [CI] 17.33–35.21); $p < 0.001$).

Incidence of intrinsic asthma (9.24% [99 of 1072] vs. 6.75% [78 of 1156]; $p = 0.03$) and extrinsic asthma (8.45% [89 of 1053] vs. 5.68% [66 of 1162]; $p = 0.01$) were higher in the *case* group. Male participants were observed more in the *case* group (68.24% [795 of 1165] vs. 43.55% [544 of 1249]; $p < 0.001$). *Case* group showed higher incidence of hay fever allergy (25.07% [272 of 1085] vs. 19.1% [224 of 1173]; $p < 0.001$) and house dust allergy (11.82%

TABLE II.
Performance Metrics.

| Metric | CV 1 | CV 2 | CV 3 | EM |
|---|---|---|---|---|
| Precision | 0.82 | 0.86 | 0.83 | 0.85 |
| Sensitivity | 0.85 | 0.77 | 0.77 | 0.85 |
| Specificity | 0.88 | 0.91 | 0.89 | 0.90 |
| F1 | 0.84 | 0.81 | 0.80 | 0.85 |
| AUROC | 0.93 | 0.92 | 0.92 | 0.95 |

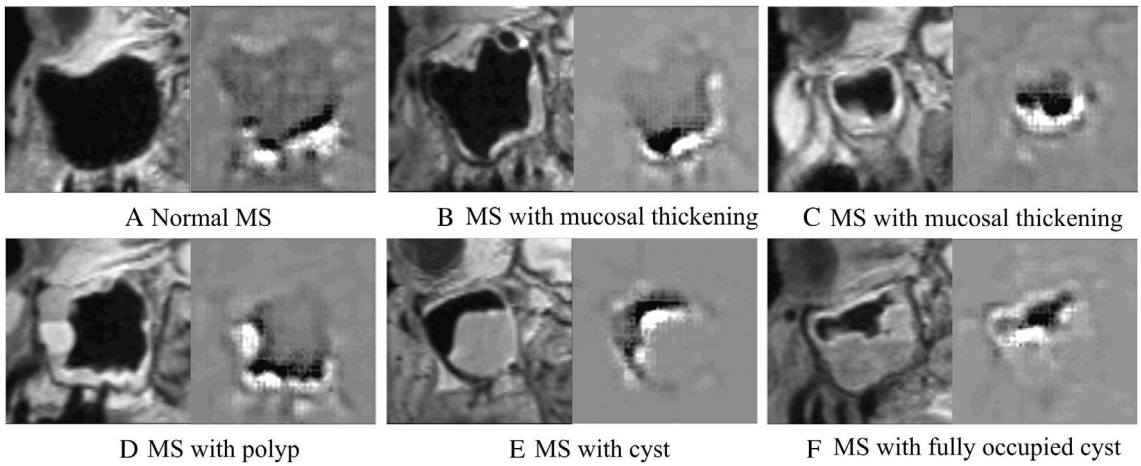


Fig. 3. Sagittal plane images and corresponding activation maps (white pixels meaning high activation and black pixels meaning no activation) of MS exhibiting no opacification, mucosal thickening, polyp, cyst, and fully occupied cyst. In. (A) Normal MS – the high activation is concentrated on the walls of the MS. (B) MS with mucosal thickening – high activation localized on the thickened mucosa. (C) MS with mucosal thickening – high activation localized on the thickened mucosa. (D) MS with polyp – activation inside the polyp mass. (E) MS with cyst – activation inside and on the edges of the cyst mass. (F) MS with fully occupied cyst – activation inside and on the edge of the fully occupied cyst mass.

TABLE III.
Comparison of Participants With No Opacification and Participants With Opacification in at Least One MS with Respect to Health and Lifestyle Factors.

| Variable | Participants with no MS opacification | Participants with MS opacification | *p*-value |
|---|---|---|---|
| Smoking habits | $N = 1245$ | $N = 1162$ | |
| Yes | 218 (17.51%) | 202 (17.38%) | 0.97 |
| No | 1027 (82.49%) | 960 (82.62%) | |
| Alcohol consumption (g/day) | $N = 1170$ | $N = 1089$ | |
| Mean (95% CI) | 15.56 (−30.18–61.31) | 20.65 (−34.83–76.15) | $6.79 \times 10^{-8}$ |
| BMI | $N = 1210$ | $N = 1126$ | |
| Mean (95% CI) | 26.26 (17.33–35.21) | 27.11 (18.99–35.33) | $3.85 \times 10^{-6}$ |
| Intrinsic asthma | $N = 1156$ | $N = 1072$ | |
| Yes | 78 (6.75%) | 99 (9.24%) | 0.03 |
| No | 1078 (93.25%) | 973 (90.76%) | |
| Extrinsic asthma | $N = 1162$ | $N = 1053$ | |
| Yes | 66 (5.68%) | 89 (8.45%) | 0.01 |
| No | 1096 (94.32%) | 964 (91.55%) | |
| Chronic bronchitis or COPD | $N = 1155$ | $N = 1069$ | |
| Yes | 67 (5.8%) | 74 (6.92%) | 0.31 |
| No | 1088 (94.2%) | 995 (93.08%) | |
| Sex | $N = 1249$ | $N = 1165$ | |
| Male | 544 (43.55%) | 795 (68.24%) | $5.5 \times 10^{-34}$ |
| Female | 705 (56.45%) | 370 (31.76%) | |
| Age (years) | $N = 1249$ | $N = 1165$ | |
| Mean (95% CI) | 63.97 (47.5–80.44) | 64.01 (47.83–80.2) | 0.90 |
| LK | $N = 1220$ | $N = 1134$ | |
| Mean (95% CI) | 6.19 (2.24–10.15) | 6.21 (2.77–9.66) | 0.76 |
| hCRP | $N = 1214$ | $N = 1126$ | |
| Mean (95% CI) | 0.22 (−0.64–1.1) | 0.23 (−0.54–1.0) | 0.91 |

[124 of 1049] vs. 8.91% [102 of 1145]; $p = 0.02$) but lower incidence of bee/wasp venom allergy (3.56% [37 of 1038] vs. 5.68% [63 of 1110]; $p = 0.02$). CI denotes confidence interval. No significant statistical correlations were found regarding smoking habits, chronic bronchitis, COPD, LK, hCRP, age, food allergy, animal hair allergy, contact allergy, medication allergy, or other allergies. Figure 4 shows the visualization of all variables with $p < 0.05$ for *control* and *case* group. We analyzed the MRI scans excluded from our study to determine the main factors behind their exclusion. Our findings indicated that borderline cases of mucosal thickening, particularly those around the 2 mm mark, exhibited lower confidence by EM. Additionally, the presence of dental accessories led to the removal of image features around the MS, thereby contributing to decreased confidence in diagnosis. Participant motion during MRI acquisition introduced noise to the data, further lowering confidence. Moreover, albeit rare, certain anatomical variations of the MS, such as those associated with Haller cells, prominent uncinate processes, hypoplasia, or surgical intervention, also contributed to diminished confidence scores. Finally, a few cases involving polyps smaller than 4 mm exhibited low confidence levels. Figure S3 shows bar chart of the image conditions which promotes low confidence score while Figure S4 shows coronal view images of maxillary sinus conditions which promote 3D CNN's low confidence scores. Figure S5 shows maxillary sinus conditions which cause 3D CNN to misclassify. For a more comprehensive understanding of our analysis and potential solutions, please refer to supplementary material section 4.

## DISCUSSION

To our knowledge, after Hansen et al. (HUNT-MRI),[5] this is the largest MRI study reporting incidental findings in the paranasal sinuses in an adult, nonselected urban population, recruited for study purposes only. Numerous studies have emphasized the importance of comprehending and addressing the prevalence of paranasal anomalies within the general population. These studies often rely on manual diagnostic methods,[1–5] where clinicians manually record opacification, requiring substantial time and effort. Subsequently, meaningful clinical hypotheses based on available participant data can be tested for statistical significance. In our work, we leverage 3D CNNs to enhance the efficiency of population studies involving large participant cohorts and use it to derive rapid clinical insights, thus reducing the workload of clinicians.

In our study, we focused on the development and validation of an EM, building upon our previous research in

TABLE IV.
Comparison of Participants With No Opacification and Participants with Opacification in at Least One MS with Respect to Different Allergies.

| Variable | Participants with no MS opacification | Participants with MS opacification | *p*-value |
|---|---|---|---|
| Hay fever | *N* = 1173 | *N* = 1085 | |
| Yes | 224 (19.1%) | 272 (25.07%) | 0.0007 |
| No | 949 (80.9%) | 813 (74.93%) | |
| Bee/wasp venom allergy | *N* = 1110 | *N* = 1038 | |
| Yes | 63 (5.68%) | 37 (3.56%) | 0.02 |
| No | 1047 (3.56%) | 1001 (96.44%) | |
| Food allergy | *N* = 1134 | *N* = 1053 | |
| Yes | 102 (8.99%) | 107 (10.16%) | 0.39 |
| No | 1032 (91.09%) | 946 (89.84%) | |
| House dust allergy | *N* = 1145 | *N* = 1049 | |
| Yes | 102 (8.91%) | 124 (11.82%) | 0.02 |
| No | 1043 (91.09%) | 925 (88.18%) | |
| Allergy to animal hair | *N* = 1152 | *N* = 1065 | |
| Yes | 85 (7.38%) | 96 (9.01%) | 0.18 |
| No | 1067 (92.62%) | 969 (90.99%) | |
| Contact allergy | *N* = 1132 | *N* = 1052 | |
| Yes | 73 (6.45%) | 49 (4.66%) | 0.08 |
| No | 1059 (93.55%) | 1003 (95.34%) | |
| Medication allergy | *N* = 1110 | *N* = 1027 | |
| Yes | 154 (13.87%) | 132 (12.85%) | 0.52 |
| No | 956 (86.13%) | 895 (87.15%) | |
| Other allergies | *N* = 1094 | *N* = 1006 | |
| Yes | 93 (8.5%) | 81 (8.05%) | 0.76 |
| No | 1001 (91.5%) | 925 (91.95%) | |

multiple instance ensembling for the classification of MS opacifications.[12] To further improve classification accuracy and prediction reliability, we employed a strategy of training three distinct CNNs on various cross-validation folds, coupled with an additional ensemble approach, a technique beneficial to DL-based CAD.[19] The robustness of our prediction model was visually demonstrated through the creation of attention maps, as seen in Figure 3. These maps revealed the CNN's focused activation on clinically significant regions during prediction processes. Notably, for cases lacking opacification, the CNN showed activation predominantly in the bottom wall, while in opacification cases, it highlighted areas such as thickened mucosal walls, polyps, and cyst masses. The precision of these attention maps played a crucial role in enhancing the reliability of our predictions, ensuring that the CNN's accuracy was not compromised by irrelevant image correlations. To show our CNN's effectiveness in deriving rapid clinical insight, we segregated participants into two categories: *control* and *case*. This separation allowed for a detailed analysis of the association between these groups and a range of clinical variables, encompassing both health and lifestyle factors as well as allergic reactions and sensitivities.

Our analysis unveiled that 48.33% (1174 of 2429) exhibited MS opacifications, comprising 13.50% (328 of 2429) with only right MS opacifications, 14.53% (353 of 2429) with only left MS opacifications and 20.29% (493 of 2429) with both left and right MS opacifications. In comparison, Hansen et al.[5], reported 66% (648 of 982) displayed opacifications in their population study. Table III presents associations related to health and lifestyle factors. Males in the *case* group displayed more MS opacifications (68.24% [795 of 1165] vs. 43.55% [544 of 1249]). This trend aligns with other studies finding similiar statistical significance with respect to sex.[1,3,5,20] The higher prevalence of MS opacifications among males may be attributable to an increased risk of allergic rhinitis.[21]

Recognizing the predominant male representation in our *case* cohort, we conducted a sex-controlled case–control analysis as elaborated in supplementary material section 5. Furthermore, participants diagnosed with bronchial asthma exhibited a heightened prevalence of MS opacifications, a trend consistent with findings reported by Hamilos et al.[22] and Zamarron et al.[23] Our study, akin to prior studies[1,3,5,20], did not find a link between smoking and MS opacifications. Our results also did not indicate a significant correlation between age and blood parameters in relation to MS opacifications. No existing literature was found correlating blood parameters with MS opacifications.

With respect to allergy related variables presented in Table IV, participants with MS opacifications also had a higher incidence of hay fever and house dust allergy. Although the literature did not show good prospective studies to address the coexistence of allergic rhinitis and rhinosinusitis (acute and chronic), many studies describe

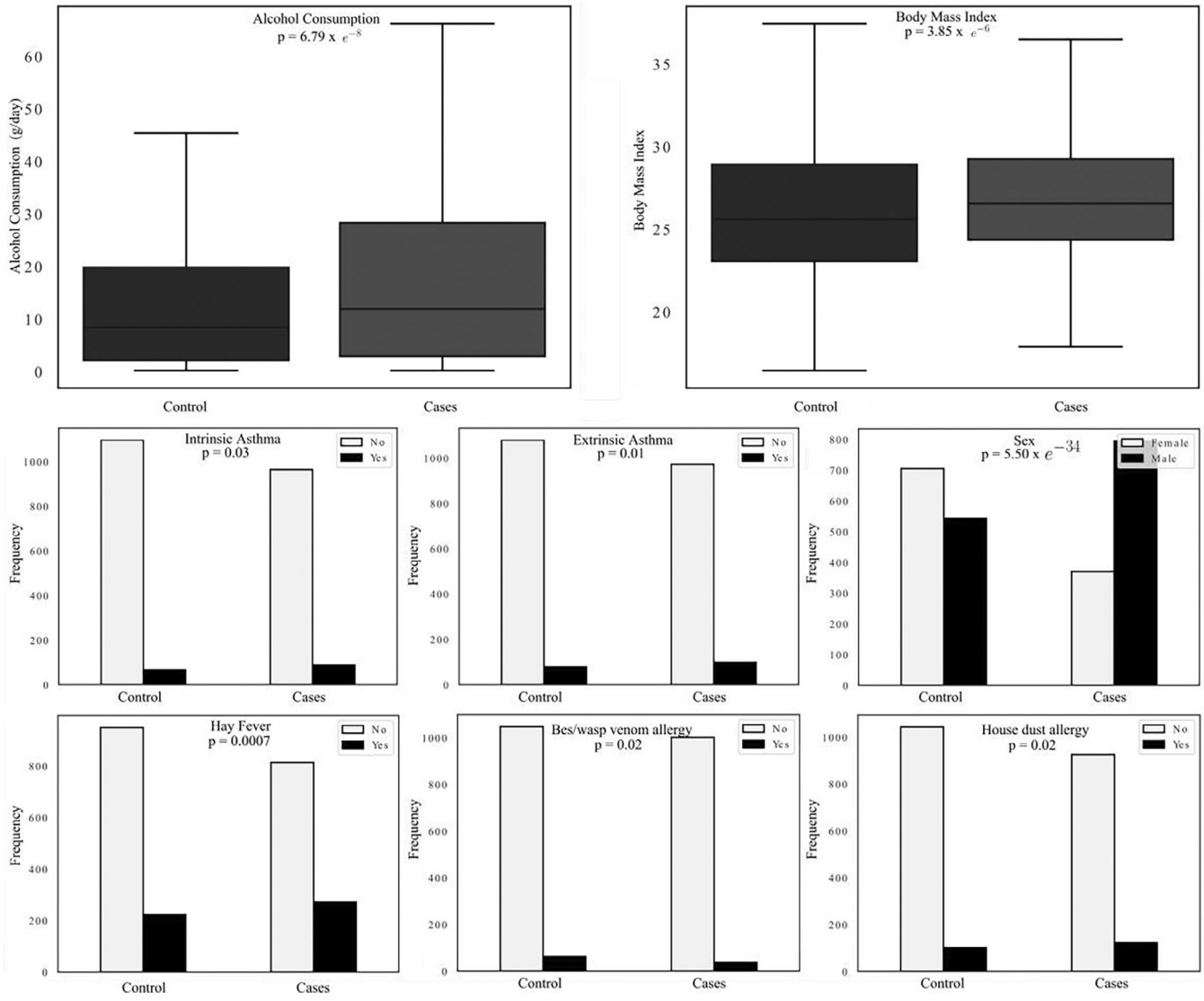


Fig. 4. Box plots and bar charts for continuous and categorical valued variables with $p < 0.05$. The variables compared are mentioned in the respective figure.

a correlation between these two entities.[24–28] Limited research has explored sinus inflammation postallergic nasal inflammation. Baroody et al.[29] and Pelikan et al.[30] demonstrated sinus inflammation through antigen-based nasal challenges. Later, Baroody et al.[31] found MS inflammation correlating with allergy seasons. Allergic rhinitis typically peaks between the second and fourth decades and diminishes thereafter,[32] notable in the context of our study's 45–74 age range participants. Slavin et al.[33] used imaging techniques of the sinuses during the ragweed season and found no changes in the sinuses.

In conclusion, our study shows methodology to develop a CNN for classifying paranasal sinuses and shows a potential application of enhancing case–control analysis in population studies. Our application reaffirms existing correlations (sex, asthma) and refutes any significant link with smoking habits, consistent with prior research. Additionally, we have uncovered previously unexplored associations (alcohol consumption, BMI, hay fever, bee/wasp venom allergy, house dust allergy) not found in the current literature. These findings underscore the potential of deep learning-based CAD, not only in enhancing MS opacifications diagnosis but also in expediting large-scale population studies. Analysis of the excluded MRIs shows the pertinent issues to consider such as noisy data, anatomically diverse MS and challenging pathological morphologies with potential techniques to redress them (see supplementary material section 4. By replacing time-consuming manual diagnosis with rapid automation, our work enables swift access to critical clinical insights. The potential for rapid acquisition of insights may enable real-time monitoring of changes across various variables as the cohort size expands, especially in a prospective study. Such capabilities are essential for effective long-term health monitoring of large populations.

## LIMITATIONS

This study has limitations. First, the training data came from one centre, potentially limiting CAD generalizability. Using multicenter, multiethnic data can enhance the reliability of our 3D CNN. Second, we only had MRI and reported clinical data. There were no data on sinonasal symptoms at the time of the MRIs, so we were unable to relate findings to current symptoms. We added questions about symptoms including the Sino-nasal Outcome test (SNOT)[34] to the protocol for the next 10000 participants, to relate the opacities specifically to sinonasal symptoms in the next evaluation. Finally, our model focuses on MS opacifications, excluding other sinuses. Future research should consider a broader sinus opacification classification.

## CONCLUSION

We present a CAD system employing CNN to classify MS opacifications and compare them with clinical data. While studies have explored prevalence of MS opacifications,[7–11] they have not been integrated into the broader context of correlating with clinical data using CAD. Our approach offers a less labor-intensive solution for detecting and classifying MS opacifications, leveraging one of the largest datasets available for studying paranasal incidental findings. We demonstrate the effectiveness of our 3D CNN model by generating attention maps, illustrating its ability to focus on clinically significant regions during the diagnostic process.


## ACKNOWLEDGMENTS

This work has not been submitted for publication anywhere else. This work is funded partially by the i3 initiative of the Hamburg University of Technology. The authors also acknowledge the partial funding by the Free and Hanseatic City of Hamburg (Interdisciplinary Graduate School) from University Medical Center Hamburg-Eppendorf. This work was partially funded by Grant Number KK5208101KS0 (Zentrales Innovationsprogramm Mittelstand, Arbeitsgemeinschaft industrieller Forschungsvereinigungen). Open Access funding enabled and organized by Projekt DEAL.


## BIBLIOGRAPHY


1. Tarp B, Fiirgaard B, Christensen T, Jensen JJ, Black FT. The prevalence and significance of incidental paranasal sinus abnormalities on MRI. *Rhinology*. 2000;38(1):33-38.
2. Rak KM, Newell JD, Yakes WF, Damiano MA, Luethke JM. Paranasal sinuses on MR images of the brain: significance of mucosal thickening. *AJR Am J Roentgenol*. 1991;156(2):381-384. https://doi.org/10.2214/ajr.156.2.1898819.
3. Cooke LD, Hadley DM. MRI of the paranasal sinuses: incidental abnormalities and their relationship to symptoms. *J Laryngol Otol*. 1991;105(4):278-281. https://doi.org/10.1017/s0022215100115609.
4. Rege ICC, Sousa TO, Leles CR, Mendonça EF. Occurrence of maxillary sinus abnormalities detected by cone beam CT in asymptomatic patients. *BMC Oral Health*. 2012;12:30. https://doi.org/10.1186/1472-6831-12-30.
5. Hansen AG, Helvik AS, Nordgård S, et al. Incidental findings in MRI of the paranasal sinuses in adults: a population-based study (HUNT MRI). *BMC Ear Nose Throat Disord*. 2014;14(1):13. https://doi.org/10.1186/1472-6815-14-13.
6. Stec N, Arje D, Moody AR, Krupinski EA, Tyrrell PN. A systematic review of fatigue in radiology: is it a problem? *Am J Roentgenol*. 2018;210(4):799-806. https://doi.org/10.2214/AJR.17.18613.
7. Kim HG, Lee KM, Kim EJ, Lee JS. Improvement diagnostic accuracy of sinusitis recognition in paranasal sinus X-ray using multiple deep learning models. *Quant Imaging Med Surg*. 2019;9:942-951. https://doi.org/10.21037/QIMS.2019.05.15.
8. Ozbay S, Tunc O. Deep Learning in Analysing Paranasal Sinuses. *Elektronika ir Elektrotechnika*. 2022;28:65-70. https://doi.org/10.5755/J02.EIE.31133.
9. Kim KS, Kim BK, Chung MJ, Cho HB, Cho BH, Jung YG. Detection of maxillary sinus fungal ball via 3-D CNN-based artificial intelligence: fully automated system and clinical validation. *PLoS One*. 2022;17(2):1-19. https://doi.org/10.1371/journal.pone.0263125.
10. Bhattacharya D, Behrendt F, Becker BT, et al. Unsupervised anomaly detection of paranasal anomalies in the maxillary sinus. In: Iftekharuddin KM, Chen W, eds. *Medical Imaging 2023: Computer-Aided Diagnosis*. Vol 12465. International Society for Optics and Photonics. SPIE; 2023:124651B.
11. Bhattacharya D, Becker BT, Behrendt F, et al. Supervised contrastive learning to classify paranasal anomalies in the maxillary sinus. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Cham: Springer Nature Switzerland; 2022:429-438.
12. Bhattacharya D, Behrendt F, Becker BT, et al. Multiple instance ensembling for paranasal anomaly classification in the maxillary sinus. *Int J Comput Assist Radiol Surg*. 2023;19:223-231.
13. Jagodzinski A, Johansen C, Koch-Gromus U, et al. Rationale and design of the Hamburg city health study. *Eur J Epidemiol*. 2019;35(2):169-181.
14. Huang G, Liu Z, Pleiss G, van der Maaten L, Weinberger K. Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell*. 2019;44:8704-8716.
15. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Institute of Electrical and Electronics Engineers Inc (IEEE); 2017.
16. Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J Open Source Softw*. 2018;3(24):638. https://doi.org/10.21105/joss.00638.
17. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17:261-272. https://doi.org/10.1038/s41592-019-0686-2.
18. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*. Institute of Electrical and Electronics Engineers Inc (IEEE); 2017:618-626.
19. Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform*. 2016;21(1):31-40.
20. Havas TE, Motbey JA, Gullane PJ. Prevalence of incidental abnormalities on computed tomographic scans of the paranasal sinuses. *Arch Otolaryngol Head Neck Surg*. 1988;114(8):856-859.
21. Tran NP, Vickery J, Blaiss MS. Management of rhinitis: allergic and nonallergic. *Allergy, Asthma Immunol Res*. 2011;3(3):148-156.
22. Hamilos DL. Chronic sinusitis. *J Allergy Clin Immunol*. 2000;106:213-227. https://doi.org/10.1067/mai.2000.109269.
23. Zamarron E, Romero D, Fern'andez-Lahera J, et al. Should we consider paranasal and chest computed tomography in severe asthma patients? *Respir Med*. 2020;169:106013. https://doi.org/10.1016/j.rmed.2020.106013.
24. Savolainen S. Allergy in patients with acute maxillary sinusitis. *Allergy*. 1989;44(2):116-122.
25. Steinke JW, Borish L. The role of allergy in chronic rhinosinusitis. *Immunol Allergy Clin N Am*. 2004;24(1):45-57.
26. Krause HF. Allergy and chronic rhinosinusitis. *Otolaryngol Head Neck Surg*. 2003;128(1):14-16.
27. Slavin RG, Spector SL, Bernstein IL, et al. The diagnosis and management of sinusitis: a practice parameter update. *J Allergy Clin Immunol*. 2005;116(6 Suppl):S13-S47.
28. Gutman M, Torres A, Keen KJ, Houser SM. Prevalence of allergy in patients with chronic rhinosinusitis. *Otolaryngol Head Neck Surg*. 2004;130(5):545-552.
29. Baroody FM, Mucha SM, Detineo M, Naclerio RM. Nasal challenge with allergen leads to maxillary sinus inflammation. *J Allergy Clin Immunol*. 2008;121(5):1126-1132.e7.
30. Pelikan Z, Pelikan-Filipek M. Role of nasal allergy in chronic maxillary sinusitis– diagnostic value of nasal challenge with allergen. *J Allergy Clin Immunol*. 1990;86(4 Pt 1):484-491.
31. Baroody FM, Mucha SM, DeTineo M, Naclerio RM. Evidence of maxillary sinus inflammation in seasonal allergic rhinitis. *Otolaryngol Head Neck Surg*. 2012;146(6):880-886.
32. Wheatley LM, Togias A. Clinical practice. Allergic rhinitis. *N Engl J Med*. 2015;372(5):456-463.
33. Slavin RG, Leipzig JR, Goodgold HM. "Allergic sinusitis" revisited. *Ann Allergy Asthma Immunol*. 2000;85(4):273-276. https://doi.org/10.1016/S1081-1206(10)62529-X.
34. Albrecht T, Beule AG, Hildenbrand T, et al. Cross-cultural adaptation and validation of the 22-item sinonasal outcome test (SNOT-22) in German-speaking patients: a prospective, multicenter cohort study. *Eur Arch Otorhinolaryngol*. 2022;279(5):2433-2439. https://doi.org/10.1007/s00405-021-07019-6.