

Daftar isi tersedia di [Sains Langsung](#)

Pengukuran

beranda jurnal: www.elsevier.com/locate/measurement

Prototipe diagnostik bantu berbasis hidung elektronik untuk deteksi kanker paru-paru dengan prediksi konformal

Xianghaozhan^{A,C}, Zhan Wang^A, Meng Yang^B, Zhi Yuan Luo^D, kamu Wang^A, Guang Li^{A,†}^ALaboratorium Kunci Negara untuk Teknologi Kontrol Industri, Institut Sistem dan Kontrol Siber, Universitas Zhejiang, Hangzhou 310027, Zhejiang, Cina^BDepartemen Ilmu dan Teknologi Komputer, Sekolah Teknik Elektronika & Informasi Mekanik, Universitas Pertambangan & Teknologi Tiongkok, Beijing 100083, Tiongkok^CDepartemen Bioteknologi, Universitas Stanford, Stanford 94305, CA, AS^DPusat Penelitian Pembelajaran Komputer, Royal Holloway, Universitas London, Egham Hill, Egham, Surrey TW20 0EX, Inggris

info artikel

Sejarah artikel:

Diterima pada 26 Agustus 2019
 Diterima dalam bentuk revisi 23 Januari 2020
 Diterima 5 Februari 2020
 Tersedia online 3 Maret 2020

Kata kunci:

Prediksi konformal
 Hidung elektronik
 Kanker paru-paru
 Prediksi daring
 Keandalan

abstrak

Kanker paru-paru menyebabkan angka kematian yang tinggi di berbagai negara sementara keandalan diagnosis kanker belum mendapat perhatian yang cukup. Dalam karya ini, aplikasi baru prediksi konformal dalam diagnosis kanker paru-paru dengan hidung elektronik diperkenalkan. Pengukuran ketidaksesuaian didasarkan pada k-tetangga terdekat. Dalam prediksi offline, akurasi sebesar 87,5% dan 83,33% telah dicapai oleh prediktor konformal berdasarkan masing-masing 1NN dan 3NN, mengungguli prediktor k-nearest neighbour yang sederhana. Selain itu, prediktor konformal memberikan informasi yang meyakinkan dan kredibel dari setiap prediksi yang dapat memberi tahu pasien tentang risiko diagnostik. Dalam prediksi online, dengan bertambahnya jumlah sampel, frekuensi kesalahan yang diberikan oleh prediksi konformal secara bertahap dapat dibatasi oleh tingkat signifikansi yang ditetapkan oleh pengguna. Proyek ini menunjukkan bahwa hidung elektronik menjanjikan alat analitik yang lebih murah dan dapat diterapkan dalam membantu diagnosis kanker paru-paru dan prediksi konformal memberikan metode yang menjanjikan untuk memastikan keandalan.

- 2020 Diterbitkan oleh Elsevier Ltd.

1. Perkenalan

Kanker paru-paru lambat laun menjadi salah satu penyakit paling mematikan dan menyumbang angka kematian yang sangat besar di berbagai negara[1–5]. Diagnosis kanker paru-paru biasanya berada pada stadium terminal dan tingkat kelangsungan hidup lima tahun cukup rendah. Diagnosis dini dapat meningkatkan tingkat kelangsungan hidup[6,7].

Para ilmuwan telah menerapkan berbagai metode untuk mendiagnosis kanker paru-paru pada tahap awal[8–11] berdasarkan biopsi. Misalnya, Tan menggunakan mesin vektor pendukung untuk menganalisis statistik ekspresi gen, dan mencapai akurasi klasifikasi sebesar 96,61%.[11]. Namun, biopsi, sebagai teknik diagnostik tercanggih dan paling andal, dapat membahayakan pasien dan tidak dapat dilakukan secara berkala dalam waktu singkat. Computed Tomography (CT) menawarkan metode non-invasif yang sering digunakan yang dapat membantu deteksi dini namun memiliki kelemahan yaitu biaya tinggi, radiasi pengion, dan hasil positif palsu. Oleh karena itu, metode diagnostik kanker paru-paru yang lebih murah, nyaman, tanpa radiasi, dan non-invasif sangat dibutuhkan.

Udara yang dihirup, sebagai hasil metabolisme yang mengandung senyawa organik yang mudah menguap (VOC), berfungsi sebagai indikator kesehatan manusia. Penggunaan udara pernafasan untuk mendiagnosis penyakit seperti kanker paru-paru telah menjadi tempat penelitian[12–14]. Misalnya, beberapa peneliti menerapkan kromatografi gas (GC) dan spektroskopi massa (MS) untuk mengetahui pola yang terkait dengan kanker paru-paru.[15,16]. Namun demikian, GC-MS cukup mahal, rumit dan memakan waktu untuk aplikasi yang luas. Hidung elektronik adalah sistem penciuman buatan yang mampu menganalisis campuran gas yang mudah menguap dengan sensor yang sensitif terhadap VOC berbeda[17–20]. Sebagai alat analisis yang relatif ringkas dan berbiaya rendah, hidung elektronik telah diterapkan pada banyak bidang seperti evaluasi kualitas lingkungan udara.[21–23], asisten diagnosis medis[24–27] dan uji mutu makanan dan minuman[28–34]. Penerapan hidung elektronik dalam diagnosis kanker paru-paru menjanjikan metode diagnostik asisten yang lebih murah dan nyaman tanpa invasi dan radiasi.

Selain itu, hal yang sering diabaikan dalam diagnosis kanker paru adalah keandalan prediksi dan akurasi keseluruhan. Interpretasi diagnostik yang salah dan prediksi yang salah dapat memberikan beban psikologis dan finansial yang sangat besar pada pasien dan keluarga mereka. Oleh karena itu, diagnosis kanker paru tidak hanya memberikan hasil prediksi tetapi juga keyakinan terkait dengan setiap prediksi, sehingga dapat membantu dokter mengambil keputusan terbaik dan memberikan informasi yang efektif kepada pasien mengenai risikonya. Untuk memecahkan

[†] Penulis yang sesuai.

Alamat email: xzhan96@stanford.edu (X.Zhan), 11732003@zju.edu.cn (Z.Wang), m.yang@cumb.edu.cn (M.Yang), zhiyuan@cs.rhul.ac.uk (Z.Luo), king_wy@zju.edu.cn (Y. Wang), guangli@zju.edu.cn (G.Li).

Masalah keandalan setiap prediksi dan keakuratan prediksi secara keseluruhan, metode seperti pembelajaran yang mungkin mendekati benar (PAC), pembelajaran Bayesian, validasi tahan dan validasi silang telah dikemukakan oleh para sarjana. Namun, PAC umumnya memerlukan sampel dalam jumlah besar dan tidak memberikan informasi spesifik mengenai keandalan prediksi individu[35]. Pembelajaran Bayesian, regresi logistik[36] dan metode Platts[37] memberikan informasi probabilistik tambahan tentang keandalan untuk setiap prediksi. Misalnya, sebelum mengeluarkan label ke suatu observasi, regresi logistik dan metode Bayesian memberikan probabilitas bersyarat berdasarkan fitur-fiturnya. Beberapa metode ini mengandalkan asumsi distribusi atau asumsi model. Misalnya, analisis diskriminan linier dan analisis diskriminan kuadrat, dua metode Bayesian, mengandalkan distribusi normal dengan kovarians yang sama atau berbeda di kelas berbeda dan menggunakan kemungkinan maksimum untuk mengevaluasi parameter distribusi. Yang lain mengandalkan data untuk menghitung apriori guna menyimpulkan distribusi kondisional sebenarnya dari setiap kelas. Karena penyimpangan sensor, gangguan pengukuran, dan kolinearitas antar fitur, asumsi distribusi dan asumsi model mungkin tidak tepat. Sementara itu, pelatihan penelitian kecil yang diadakan pada penelitian biomedis biasanya bersifat bias di kelas. Frekuensi kasus penyakit mungkin bukan merupakan perkiraan akurat dari probabilitas sebelumnya. Oleh karena itu, model dan keandalannya mungkin tidak cukup kuat dalam proses pencarian informasi sebelumnya. Selain itu, metode seperti validasi penundaan dan validasi silang sering kali memprediksi kenyataan secara berlebihan karena respons yang diberikan oleh hidung elektronik mungkin dipengaruhi oleh penyimpangan sensor, penuaan sensor, atau keracunan sensor.

Prediksi konformal, yang didasarkan pada asumsi bahwa semua sampel dan label asosiasinya dihasilkan dari distribusi independen dan identik (IID), pertama kali dikemukakan oleh Vladimir Vovk. dan rekan kerjanya[35,36,38,39]. Berdasarkan asumsi yang lebih lemah jika dibandingkan dengan metode yang disebutkan di atas, bila diterapkan pada data aktual dari hidung elektronik, metode ini dapat memberikan metode yang menjanjikan untuk evaluasi keandalan prediksi. Prediksi konformal tidak hanya memberikan prediksi individu, tetapi juga mengungkapkan kesesuaian setiap prediksi baru jika dibandingkan dengan kelompok pengamatan yang sudah diperiksa, mampu menawarkan informasi tambahan namun penting tentang kepercayaan dan kredibilitas setiap prediksi dan menghindari penilaian berlebihan terhadap keakuratan prediksi secara keseluruhan.[40].

Dalam penelitian ini, sampel udara pernapasan dari pasien kanker paru-paru dan kontrol dikumpulkan dan dianalisis dengan hidung elektronik, yang diikuti dengan penggunaan prediksi konformal berdasarkan k-tetangga terdekat untuk klasifikasi sampel kanker paru-paru. Di bagian2, definisi prediksi konformal dan penerapannya dalam prediksi online dan prediksi offline diilustrasikan masing-masing. Kemudian, proses pengambilan sampel, sistem hidung elektronik dan metode pengolahan data diilustrasikan pada Bagian3. Hasil dan implikasi dibahas dan dianalisis di Bagian4. Akhirnya diambil kesimpulan.

2. Metode

2.1. Prediksi konformal

2.1.1. Definisi

Dalam masalah klasifikasi pembelajaran mesin, biasanya terdapat sekumpulan sampel sebagai set pelatihan.

Setiap pengamatan berisi satu objek $X_{\text{Saya}2X}$ dan satu label $k_{\text{kamu}2\text{Saya}2}$ kamu, di mana X menunjukkan ruang objek dan Y menunjukkan ruang label. Selama sampel baru $X_{\text{ndiberikan}}$, tugas klasifikasi adalah memprediksi label milik sampel tersebut.

$$DDX_i; k_{\text{kamu}1TH}; \dots; DX_{N-1}; k_{\text{kamu}N-1HAI}$$

$$Z_{\text{Saya}14}DX_{\text{Saya}1\text{kamu}2\text{Saya}1TH}; S_{\text{Saya}141}; 2; \dots; DN-1TH$$

Dengan menggabungkan sampel dan label, ruang contoh Z dapat diatur. Apa yang dilakukan prediktor sederhana adalah menemukan satu pemetaan F yang membantu prediksi sampel baru $X_{\text{ndiberdasarkan}}$ ruang contoh saat ini Z .

$$F: Z \rightarrow Y$$

D3TH

Jika dibandingkan dengan prediktor sederhana, prediktor konformal memiliki parameter lain-280; 1TH, yang didefinisikan sebagai tingkat signifikansi. Sementara itu, 1--didefinisikan sebagai tingkat kepercayaan, yang mencerminkan keyakinan prediksi individu. Prediktor konformal mengeluarkan subset ruang label Y berdasarkan tingkat signifikansi tertentu:

$$C-DDX_i; k_{\text{kamu}1TH}; \dots; DX_{N-1}; k_{\text{kamu}N-1TH}; X_{\text{ndiberdasarkan}}$$

D4TH

Output himpunan bagian oleh prediktor konformal disarangkan, seperti yang ditunjukkan di bawah ini:

$$C-1DZ_1; \dots; Z_{N-1}TH; X_{\text{ndiberdasarkan}} C-2DZ_1; \dots; Z_{N-1}TH; X_{\text{ndiberdasarkan}} P-2TH$$

D5TH

Untuk menghasilkan kumpulan prediksi, perlu diperkenalkan fungsi terukur A yang memetakan setiap observasi $z_{\text{Saya}14}D_{\text{Saya}141}; 2; \dots; N_{\text{P}}2Z_{\text{ke}}$ bilangan real $A_{\text{Saya}14}D_{\text{Saya}141}; 2; \dots; N_{\text{P}}2R$. Proses tersebut didefinisikan sebagai pengukuran ketidaksesuaian, yang mencerminkan betapa tidak lazimnya setiap pengamatan ketika ditempatkan dalam kelompok pengamatan lainnya:

$$A_{\text{Saya}14}ADZ_1; \dots; Z_{\text{Saya}14}S_{\text{Saya}141}; \dots; Z_{\text{NTH}}; S_{\text{Saya}141}; \dots; N$$

D6TH

Selain itu, A_{harus} didasarkan pada algoritma pembelajaran statistik tertentu dan memenuhi properti asumsi pertukaran berikut: untuk n dan permutasi apa pun P ,

$$DA_1; \dots; A_{N_{\text{P}}}ADZ_1; \dots; Z_{N_{\text{P}}-1}AD_{\text{PD}1TH}; \dots; A_{\text{PD}NTH}TH$$

$$1/4ADZ_{\text{PD}1TH}; \dots; Z_{\text{PD}NTH}TH$$

D7TH

Kemudian, prediktor konformal yang bergantung pada A didefinisikan sebagai:

$$C-DZ_1; Z_2; \dots; Z_{N-2}; Z_{N-1}; Z_{N_{\text{P}}}1/4fk_{\text{kamu}1\text{P}}>-G$$

D8TH

Untuk sampel baru $X_{\text{ndiberdasarkan}}$, prediktor konformal mengeluarkan semua kemungkinan label dalam kumpulan C -berdasarkan tingkat signifikansi tertentu - 280; 1TH dan set pelatihan. Untuk setiap kemungkinan label $k_{\text{kamu}2\text{kamu}}$, nilai p yang terkait dengannya didefinisikan sebagai:

$$P_{\text{kamu}14} \frac{jfS_{\text{Saya}141}; \dots; N_{\text{JA}}k_{\text{kamu}2\text{Saya}1}A_{\text{ndiberdasarkan}}}{N}$$

D9TH

Di mana P_{kamu} mewakili bahwa ketika label $X_{\text{ndiberdasarkan}}$ adalah y , seberapa baik observasi yang tidak terlihat tersebut sesuai dengan observasi lainnya.

Urutan skor ketidaksesuaian yang sesuai untuk satu label prediksi y ditentukan oleh:

$$DA_{\text{ndiberdasarkan}}A_{\text{kamu}2}; \dots; A_{\text{kamu}N_{\text{P}}}1/4ADZ_1; Z_2; \dots; Z_{N-1}; DX_{\text{ndiberdasarkan}}HAI$$

D10TH

Berdasarkan $A_{\text{kamu}N}$, diakui semakin rendah A_{kamu} adalah, semakin tinggi keyakinan yang kami miliki terhadap prediksi tersebut, karena ini berarti kombinasi fitur dan label ini lebih sesuai dengan pengamatan lainnya.

Untuk prediksi konformal, validitas prediksi berarti:

$$PD_{\text{kamu}N}RC-DDX_i; k_{\text{kamu}1TH}; \dots; DX_{N-1}; k_{\text{kamu}N-1TH}; X_{\text{ndiberdasarkan}}HAI6-$$

D11TH

2.1.2. Pengukuran ketidaksesuaian

Secara teoritis, algoritme apa pun dapat dimodifikasi menjadi algoritme dasar yang menjadi dasar pengukuran ketidaksesuaian. Vladimir Vovk. awalnya mendasarkan perhitungan ketidaksesuaian prediksi konformal pada k-tetangga terdekat. Metode khas berdasarkan k-tetangga terdekat (KNN) juga diterapkan dalam penelitian ini. Untuk membuatnya sederhana dan jelas, CP-1NN dan CP-3NN digunakan untuk menunjukkan prediktor konformal berdasarkan masing-masing 1NN dan 3NN dan prediktor sederhana yang sesuai dilambangkan sebagai 1NN dan 3NN. Algoritma pengukuran ketidaksesuaian diilustrasikan seperti di bawah ini:

D1TH

D2TH

Untuk pengamatan tertentu $DX_{\text{Saya}}; \text{kamu}_{\text{Saya}}P, \text{yaPertama}$, jarak antara observasi ini dan observasi lain di set pelatihan dihitung, yang dinotasikan sebagai:

$$DDX_{\text{Saya};X_jTH;j\frac{1}{4}1;2;\dots; \text{Saya}-1;\text{Sayath}1;\dots;N$$

Kemudian, k observasi terdekat dengan label yang sama dengan $DX_{saya; kamu; s; TH}$ ditemukan dan dilambangkan sebagai $DX_{adalah; kamu; adalah; TH, s = 1, \dots, k}$. Sedangkan k observasi terdekat dengan label yang berbeda dengan $DX_{saya; kamu; s; TH}$ ditemukan dan dilambangkan sebagai $DX_{is; kamu; is; TH, s = 1, \dots, k}$. Skor ketidaksesuaian diberikan oleh jumlah jarak antara observasi tes dan k observasi homogen terdekat (dengan label yang sama) dibagi dengan jumlah jarak antara observasi tes dan k observasi heterogen terdekat (dengan label berbeda). K adalah kumpulan angka untuk observasi heterogen dan observasi homogen.

$$A_{\text{saya}} \frac{1}{4} P_k \frac{5 \text{ mDDX}_{\text{sayu}} \text{kamu}_{\text{Lation}} \text{TH}}{5 \text{ mDDX}_{\text{is}} \text{kamu}_{\text{is}} \text{TH}} P_k$$

Berdasarkan metode yang tercantum di atas, semakin dekat pengamatan baru dengan label yang sama, semakin baik kesesuaian pengamatan baru tersebut dan semakin rendah Adalah.

2.1.3. Prediksi dalam mode online

Sebagai salah satu protokol yang paling umum digunakan untuk masalah pembelajaran mesin, prediksi online menerapkan prinsip langkah demi langkah yang praktis dan berguna dalam penerapan di dunia nyata. Dalam mode ini, setelah sejumlah pengamatan tetap dicantumkan dalam set pelatihan, labelnya y_{n+1} dari sampel baru x_{n+1} diprediksi berdasarkan set pelatihan yang ada:

DDX₁;kamu₁TH; DX₂;kamu₂TH; . . . ; DX_N;kamu_NHAI D14TH

Kemudian efektivitas prediksi dievaluasi dari perspektif prediksi yang salah/berbuat salah- N , beberapa prediksi banyak/atau prediksi kosong/kosong- N , yang akan diilustrasikan nanti. Setelah itu, observasi baru dengan label yang benar ini, ditambahkan ke set pelatihan untuk memperbesar jumlah observasi yang tersedia untuk prediksi lebih lanjut. Prosesnya dilakukan berulang kali, artinya set pelatihan terus diperbarui selama seluruh proses prediksi online. Langkah-langkah yang menunjukkan cara kerja mode online tercantum sebagai berikut[40]:

PROTOKOL PREDIKSI ONLINE:

Berikut ini hasil uji t pada variabel-variabel yang diteliti:

Keluaran prediktorC-_NYuntuk semua-2đ0; 1TH;
Keluaran realitaskamun2kamu;

berbuat_N $\frac{1}{4}$ f₁ 0 jika kamu_{RC}
jika tidak
Berbuat_{RC} berbuat salah_N $\frac{1}{4}$ berbuat salah_N
banyak_N $\frac{1}{4}$ f₁ 0 jika_C $\frac{1}{4}$ > 1
jika tidak
Banyak_N $\frac{1}{4}$ Banyak_N $\frac{1}{4}$ t_h banyak_N
kosong_N $\frac{1}{4}$ f₁ 0 jika_C $\frac{1}{4}$ $\frac{1}{4}$ > 0
jika tidak
Em- $\frac{1}{4}$ Em- $\frac{1}{4}$ t_h kosong_N
Set pelatihan = Set pelatihan [DX]
AKHIR UNTUK

Kapan rumus(11)dan hukum kuat dalam jumlah besar dipertimbangkan, validitas prediktor konformal online ditunjukkan bahwa:

batassup $\frac{\text{Berbuat salah.}}{N}$ N6-

D15TH

2.1.4. Prediksi dalam mode offline

Prediksi offline adalah kebalikan dari prediksi online dan ditandai dengan prediksi berdasarkan kumpulan pelatihan tetap dan model tetap. Dalam proyek ini, prediksi offline dilakukan dalam mode validasi silang keluar-keluar. Prediksi dalam mode offline bergantung pada aturan tertentu yang diperoleh dari set pelatihan statis, dan validitas prediksi konformal dalam mode offline tidak terbukti secara ketat. Prediksi konformal dalam mode offline dapat memberi pengguna informasi tambahan mengenai keyakinan untuk setiap prediksi, bukan hanya memberikan prediksi. Informasi keandalan memberikan informasi risiko kepada pengguna untuk mengambil keputusan yang lebih baik. Untuk prediksi setiap individu, prediktor konformal dapat mengeluarkan label dengan nilai p tertinggi, yang didefinisikan sebagai prediksi paksa. Bersamaan dengan itu, prediktor konformal menawarkan informasi keandalan dengan dua fitur, kepercayaan dan kredibilitas:

kepercayaan diri¹⁴supF1 --:JC-J61G

kredibilitas¼infF-;jC-j¼0G

Dalam soal klasifikasi, keyakinan sama dengan 1 dikurangi nilai p terbesar kedua, yang mewakili keyakinan prediksi ketika label prediksi dengan nilai p terbesar dipilih sementara label prediksi lainnya ditolak. Kredibilitas sama dengan nilai p terbesar, yang menunjukkan seberapa baik kesesuaian pilihan yang dipilih.

Suatu prediksi dinilai dapat diandalkan selama keyakinannya mendekati 1 dan kredibilitasnya tidak mendekati 0 [38], artinya tidak ada pilihan yang lebih baik jika dibandingkan dengan prediksi ini. Jika kredibilitas mendekati 0, prediksi tersebut cenderung tidak konformal dalam kelompok dan mungkin merupakan outlier yang tidak dapat diandalkan.

2.2. Eksperimen dan pemrosesan data

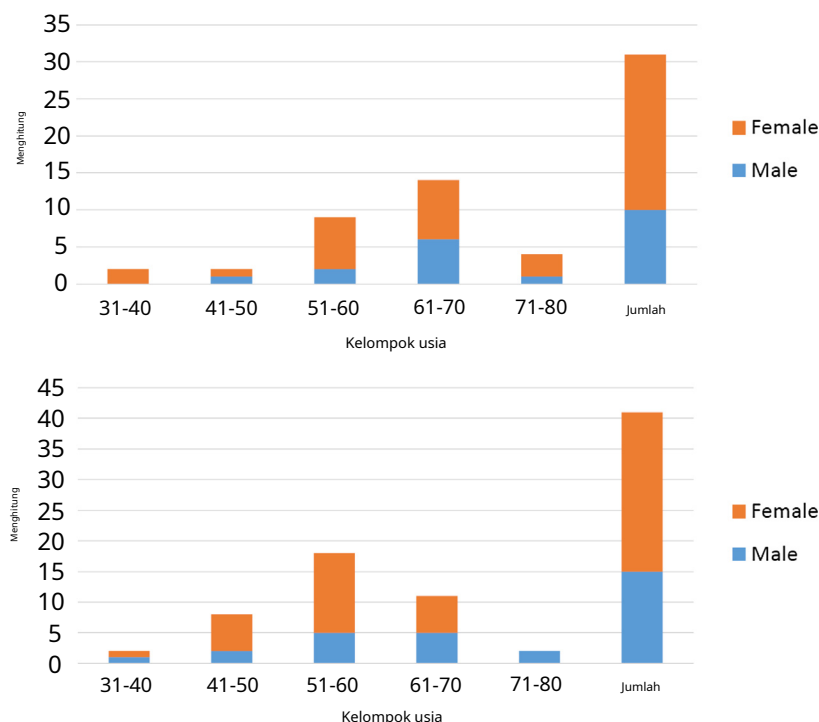
2.2.1. Pengumpulan sampel

Sampel dikumpulkan dari pasien berusia 30 hingga 80 tahun yang pertama kali didiagnosis menderita kanker paru-paru dari Rumah Sakit Afiliasi Kedua Universitas Zhejiang. Sementara itu, pasien yang dipilih belum menerima kemoterapi atau radioterapi selama 6 bulan sebelum tes dan tidak merokok setidaknya selama 6 bulan. Selain itu, sampel kontrol berasal dari guru dan pejabat yang memiliki rasio gender dan sebaran usia yang sama dengan pasien. Mereka berasal dari Universitas Zhejiang dan tidak didiagnosis menderita penyakit pernapasan, diabetes, atau masalah kesehatan lainnya yang mungkin berdampak negatif pada eksperimen tersebut. Sementara itu, mereka belum menjalani operasi apa pun dalam satu tahun. Semua sukarelawan yang diuji menawarkan persetujuan tertulis. Distribusi usia sukarelawan ditunjukkan pada [Gambar 1](#).

Di antara pasien tersebut, 26 kasus merupakan kasus adenokarsinoma, 1 kasus karsinoma sel kecil, 2 kasus karsinoma sel skuamosa, dan 2 kasus karsinoma sel besar. 13 kasus terjadi di paru kiri pasien, sedangkan 18 kasus terjadi di paru kanan.

Semua sukarelawan dilarang makan setelah pukul 22:00 sehari sebelum pengambilan sampel. Kemudian sampel diambil pada pukul 06.00 hingga 07.00 pagi dengan perut kosong. Relawan bisa mencuci mulut dengan air tapi tidak bisa menyikat gigi. Sementara itu, mereka dilarang melakukan aktivitas berat seperti berlari selama 2 jam sebelum ujian. Semua sampel dikumpulkan di bawah instruksi peneliti.

Peralatan pengumpulan udara offline M3014-4 (ECO MEDICS AG, Duernten, Swiss) digunakan untuk proses pengambilan sampel. Kantung udara terbuat dari polietilen tereftalat aluminisasi. Ilustrasi peralatan pengambilan sampel ditunjukkan pada Gambar 2. Pertama, seorang sukarelawan menarik napas melalui mulut. Seperti yang ditunjukkan oleh

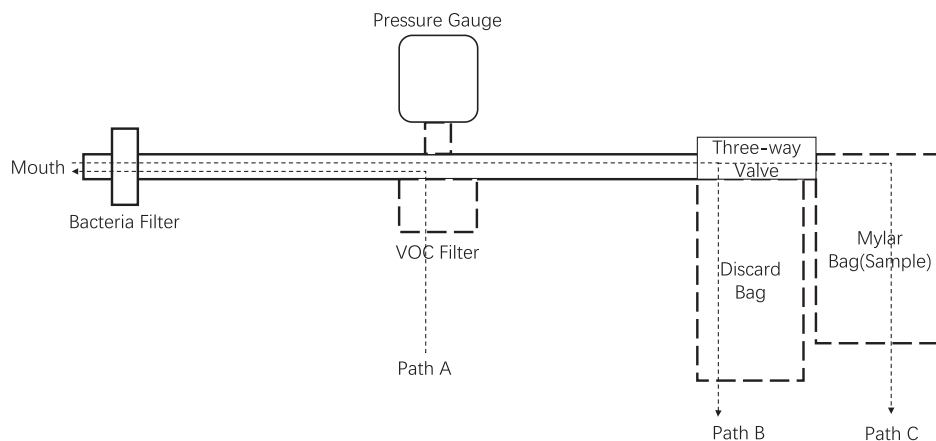


Gambar 1. Sebaran usia dan jenis kelamin relawan pada kelompok pasien (atas) dan kelompok kontrol (bawah).

Jalur A, udara melewati filter VOC dan filter bakteri lalu masuk ke dalam mulut. Prosesnya berlangsung selama 3–5 menit. Saat udara disaring, proses ini dimaksudkan untuk mengurangi dampak lingkungan. Kedua, seorang relawan menarik napas dalam-dalam dan menghembuskannya ke dalam mulut. Persyaratannya adalah tekanan harus dipertahankan dalam kisaran 10 - 15 cmH₂O untuk mengurangi pengaruh laju aliran. Bagian pertama dari udara pernafasan dikumpulkan oleh Kantong Buang melalui Jalur B dan saat Kantong Buang terisi, udara pernafasan mengikuti Jalur C ke dalam Kantong Mylar. Proses ini memastikan sampel udara napas mewakili udara di dalam paru-paru dan mengurangi pengaruh udara di trakea dan rongga mulut. Terakhir, sampel disimpan dalam termostat pada suhu 37 derajat Celcius, diangkut ke laboratorium dan dianalisis dengan hidung elektronik dalam waktu satu jam. Sebanyak 72 sampel udara pernafasan (termasuk 31 sampel dari pasien dan 41 sampel dari kontrol) dikumpulkan dengan pengumpul udara.

2.2.2. Sistem dan eksperimen hidung elektronik

Semua percobaan dan analisis didasarkan pada hidung elektronik yang dibuat di Laboratorium Kunci Negara Teknologi Kontrol Industri di Universitas Zhejiang [40–43], dengan 16 sensor semikonduktif oksida logam (MOS), tipe TGS dan dibeli dari Figaro Engineering Inc. (Osaka, Jepang). Sensor jenis ini bekerja dengan baik dalam tugas klasifikasi multipel [44,45]. Ke-16 sensor memiliki spesifisitas yang tumpang tindih [42,40]. Dan keandalan serta keakuratan sistem hidung elektronik ini telah diverifikasi dalam penelitian sebelumnya [42,40]. Misalnya, kita [46] telah berhasil menggunakan sistem ini untuk mengklasifikasikan 12 kategori obat herbal alternatif yang berbeda dengan akurasi validasi silang satu kali keluar (LOOCV) sebesar 98,94% selama 600 observasi. Terlebih lagi, kami juga telah memvalidasi efektivitas pembelajaran online dengan electronic nose [47]. Dalam karya lain, kami [48] menerapkan hidung elektronik untuk membedakan asal usul yang berbeda dari jenis herbal alternatif yang sama.



Gambar 2. Ilustrasi peralatan dan proses pengambilan sampel.

kedokteran dan mencapai akurasi LOOCV berkisar antara 85,63% hingga 99,78% untuk setiap tugas klasifikasi.

Deskripsi skema sistem e-nose dapat direpresentasikan dalam Gambar 3, yang mencakup susunan sensor dalam kamar gas 200 ml, katup tiga arah yang berubah dari aliran sampel dan udara bersih standar, pompa udara yang memompa udara dengan kecepatan 1 L/mnt, sistem catu daya, dan unit akuisisi data. Katup tiga arah adalah pengontrol utama gas uji dan udara bersih standar. Ini adalah hidung elektronik umum yang digunakan untuk aplikasi multifungsi [40,41]. Komponen sensitif khas dari setiap sensor ditunjukkan pada Tabel 1 tetapi sensornya tidak hanya sensitif terhadap komponen yang terdaftar. Sensor tersebut mampu bereaksi terhadap senyawa organik yang mudah menguap dari udara yang dihirup dengan beragam komponen termasuk fenol dan aldehida. Sementara itu, mereka tidak terlalu sensitif terhadap satu jenis gas tertentu, sehingga sesuai untuk analisis udara pernapasan yang melibatkan senyawa organik kompleks yang mudah menguap.

Seluruh proses percobaan pengukuran setiap sampel yang dikumpulkan diilustrasikan dalam Gambar 4 [40,46–48]: 400 detik ditetapkan untuk setiap percobaan sampel uji individu. Frekuensi pengambilan sampel sistem hidung elektronik adalah 100 Hz. Pertama-tama, udara bersih standar dipompa ke panel sensor dengan kecepatan 1 L/mnt selama 20 detik agar respons sensor kembali ke kondisi awal. Kemudian, pada saat aliran udara bersih standar dihentikan, sampel gas uji diinjeksikan ke dalam ruangan sesegera mungkin. Injektor medis digunakan untuk mengekstraksi campuran gas 10 ml. Setelah itu, kami menetapkan 180 detik untuk periode reaksi dan mencatat respons sensor dalam bentuk perubahan tegangan. Pada $t = 200$ detik, udara bersih standar dipompa lagi ke dalam sistem dan sampel gas uji dipompa keluar. Dari $t = 200$ detik hingga $t = 340$ detik, kami mencatat pola penurunan respons sensor. Akhirnya, 60 detik lagi ditetapkan agar sensor lebih stabil dan kembali ke garis dasar.

Setelah mengambil sampel udara pernapasan, semua pengujian dengan e-nose dilakukan di laboratorium yang sama dengan hidung elektronik yang sama pada suhu lingkungan yang bervariasi antara 22 hingga 25 derajat Celsius dan dengan catatan kelembaban relatif berkisar antara 50 hingga 65%.

2.2.3. Pemrosesan data dan ekstraksi fitur

Pemrosesan data pekerjaan ini dilakukan pada komputer pribadi tanpa akselerasi GPU. Respons sensor yang khas ditunjukkan pada Gambar 5, yang menunjukkan respons terhadap sampel kontrol pertama dengan setiap kurva mewakili salah satu dari 16 sensor. Waktu pengambilan sampel adalah 0,01 detik. Transformasi Wavelet Diskrit (db2) digunakan untuk menguraikan sinyal dan tingkat kelima dari sinyal terurai dipilih untuk menangani kebisingan. Kemudian, semua sinyal yang terekam dikalibrasi dengan mengurangi nilai dasar untuk meminimalkan pengaruh penyimpangan sensor:

$$V_{\text{norm}} = V - V_0$$

D18TH

V_0 dilambangkan respon sensor setelah penyaringan dan V_0 menunjukkan nilai dasar setelah pemfilteran.

Tabel 1

Informasi sensor yang digunakan dalam susunan 16 sensor di hidung elektroninc.

Sensor Jenis	Menghitung	Paling sensitif terhadap:
TGS800	1	Karbon monoksida, etanol, metana, hidrogen, amonia
TGS813	2	Karbon monoksida, etanol, metana, hidrogen, isobutana
TGS816	1	Karbon monoksida, etanol, metana, hidrogen, isobutana
TGS821	1	Karbon monoksida, etanol, metana, hidrogen
TGS822	2	Karbon monoksida, etanol, metana, aseton, n-heksana, benzena, isobutana
TGS826	1	Amonia, trimetil amina Etanol,
TGS830	1	R-12, R-11, R-22, R-113
TGS832	1	R-134a, R-12, R-22, etanol Karbon
TGS880	1	monoksida, etanol, metana, hidrogen, isobutana
TGS2620	1	Karbon monoksida, metana, isobutana, hidrogen
TGS2600	1	Karbon monoksida, hidrogen Hidrogen,
TGS2602	1	amonia etanol, hidrogen sulfida, toluena
TGS2610	1	Etanol, hidrogen, metana, isobutana, propana
TGS2611	1	Etanol, hidrogen, isobutana, metana

Setelah itu, berdasarkan kinerja penambangan data deret waktu sebelumnya [40,41,46–48], 9 fitur yang umum digunakan untuk e-nose (total $9 \times 16 = 144$ fitur) diekstraksi:

1. Nilai Maksimum

$$V_{\text{maks}} = \max(V)$$

D19TH

2. Nilai Integral

$$Z_T = \int_0^T V_{\text{norm}} dt$$

D20TH

T melambangkan seluruh waktu pengukuran, $T = 340$ detik.

3. Fase Informasi

$$Z_{\text{JVM}} = \int_0^T V_{\text{norm}} \frac{dV_{\text{norm}}}{dt} dt$$

D21TH

4–9. Rata-rata pergerakan eksponensial dari turunan V

$$E_{\text{DVP}} = \frac{1}{T} \int_0^T V_{\text{norm}} e^{-k(t-t_0)} dt$$

D22TH

Rata-rata pergerakan eksponensial pengambilan sampel diskrit k_{TH} dan faktor penghalusan A didefinisikan sebagai:

$$k_{\text{TH}} = \frac{1}{T} \ln \left(\frac{V_{\text{norm}}(t_0) - V_{\text{norm}}(t_0 + T)}{V_{\text{norm}}(t_0) - V_{\text{norm}}(t_0 + 1)} \right)$$

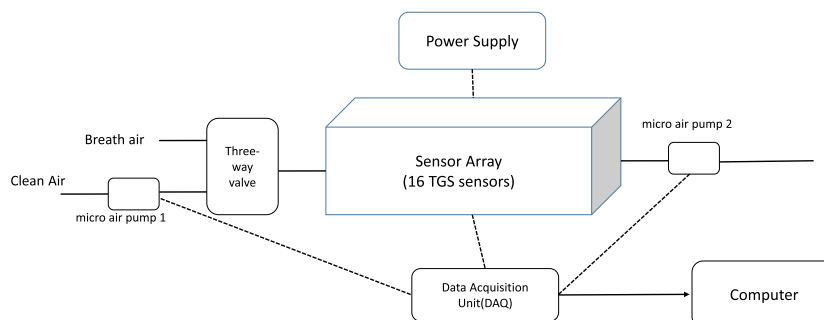
D23TH

$$A = \frac{1}{100 - SR} \cdot \frac{1}{10 - SR} \cdot \frac{1}{SR}$$

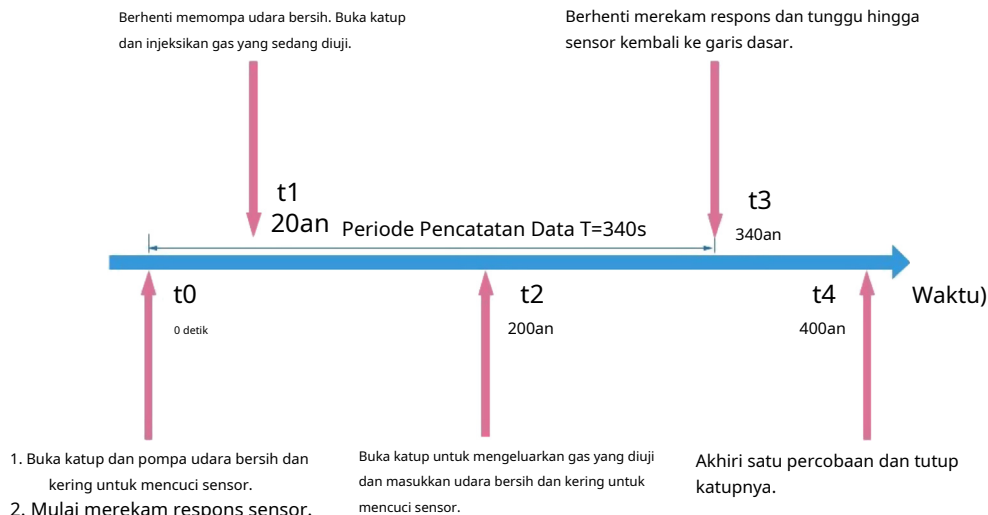
D24TH

$$k_{\text{TH}} = \frac{1}{T} \ln \left(\frac{V_{\text{norm}}(t_0) - V_{\text{norm}}(t_0 + T)}{V_{\text{norm}}(t_0) - V_{\text{norm}}(t_0 + 1)} \right)$$

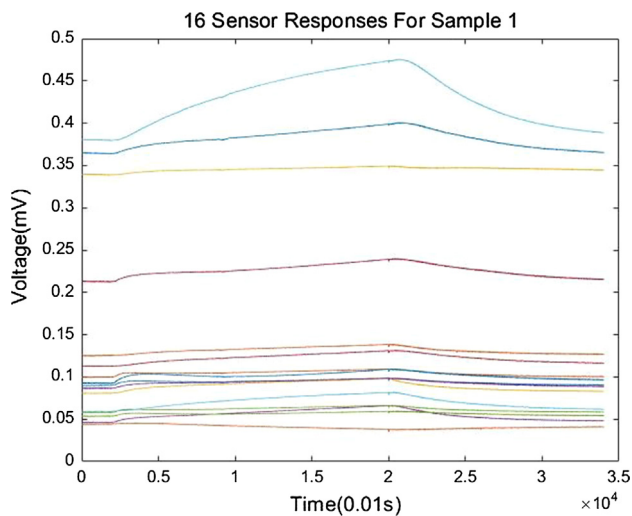
D25TH



Gambar 3. Ilustrasi singkat sistem hidung elektronik yang digunakan dalam proyek ini.



Gambar 4.Seluruh proses dalam satu percobaan.



Gambar 5.Kurva respons sensor tipikal yang diberikan oleh sistem e-nose.

label dengan nilai p tertinggi adalah keluarannya, dan hasil prediksi sederhana yang diberikan oleh 1NN dan 3NN disajikan dalam [Meja 2](#).

Berdasarkan hasil, dalam hal keakuratan prediksi, kinerja CP-1NN sama baiknya dengan 1NN sedangkan CP-3NN mengungguli 3NN. Tanpa mengorbankan keakuratan, prediksi konformal memberikan informasi tambahan tentang keyakinan dan kredibilitas yang mendasari setiap prediksi. Untuk lebih konkretnya, lima prediksi tipikal dipilih secara acak dan hasilnya ditampilkan [Tabel 3](#)(CP-1NN). Label 0 menunjukkan kontrol sedangkan label 1 menunjukkan pasien. Berdasarkan hasil CP-1NN yang diberikan oleh prediktor paksa, sampel pertama diprediksi berlabel 0 dengan keyakinan 0,9167 yang menunjukkan bahwa prediktor tersebut yakin akan menolak label lainnya. Sedangkan prediksi ini memiliki kredibilitas sebesar 0,1667 yang menunjukkan bahwa meskipun label lain tidak sesuai dibandingkan dengan label yang dipilih, namun prediksi tersebut sendiri tidak terlalu sesuai dengan set pelatihan dan keandalan prediksi tersebut tidak tinggi. Sampel ke-59 diprediksi memiliki label 1 yang benar, dengan keyakinan 0,9861 dan kredibilitas 0,9306, yang menunjukkan bahwa prediksi ini dapat diandalkan karena tidak hanya sesuai dengan set pelatihan tetapi juga lebih baik daripada label rekanannya.

Untuk sampel 1 dan sampel 64, kredibilitas yang lebih rendah berarti hasilnya kurang dapat diandalkan. Dalam keadaan ini ketika keandalan prediksi tidak cukup tinggi, meskipun diagnosis spesifik dapat diberikan, dokter harus memberi tahu pelanggan tentang risiko diagnosis yang menyesatkan dan merekomendasikan pelanggan untuk terus memantau kesehatan mereka dan melakukan analisis medis lebih lanjut jika tingkat kepercayaan diri rendah. atau kredibilitas rendah terjadi, karena diagnosis yang salah dapat menimbulkan beban besar bagi pelanggan.

Dari hasil dalam mode offline, selain memberikan label prediksi, prediktor konformal memberikan informasi reliabilitas tambahan dan signifikan tentang prediksi individu dengan keyakinan dan kredibilitas, yang mungkin memainkan peran penting dalam diagnosis.

SR $\frac{1}{4}$ 100 menunjukkan tingkat pengambilan sampel.EaDVTH dilambangkan dengan vektor yang memuat nilai terbesar dan nilai terkecil dalam kurun waktu setelah injeksi sampel gas. Ekstraksi fitur dipandu oleh penelitian sebelumnya tetapi evaluasi signifikansi fitur, optimalisasi sensor, dan informasi yang lebih efektif dapat dilakukan, yang terbuka untuk penelitian lebih lanjut di masa mendatang. Setelah itu, fitur-fitur ini digabungkan menjadi 72/144 matriks dengan setiap kolom mewakili satu fitur. Terakhir, setiap kolom fitur dinormalisasi dengan pengurangan minimum dan pembagian rentang setiap fitur. Setelah normalisasi, fitur berada pada rentang 0 hingga 1. Semua prapemrosesan data dan pelatihan model dilakukan dengan MATLAB R2017b.

3. Hasil dan Pembahasan

3.1. Prediksi offline dengan prediksi konformal

Dalam mode prediksi offline, protokol validasi silang satu kali keluar diterapkan. Hasil prediksi paksa yang diberikan oleh prediktor konformal dengan 1NN (CP-1NN) dan 3NN (CP-3NN), dimana

Meja 2

Akurasi Prediksi Konformal Dalam Mode Offline.

Metode Prediksi	1NN	3NN
Prediksi Konformal yang Dipaksa	87,50%	83,33%
Prediksi Sederhana	87,50%	81,94%

Tabel 3

Lima prediksi individu tipikal dengan CP-1NN.

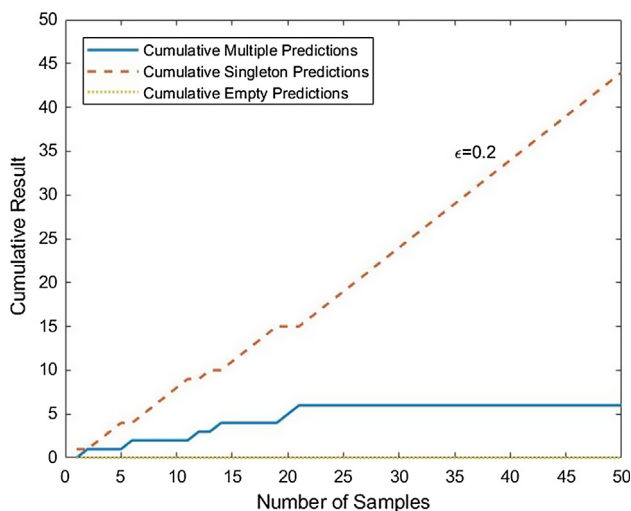
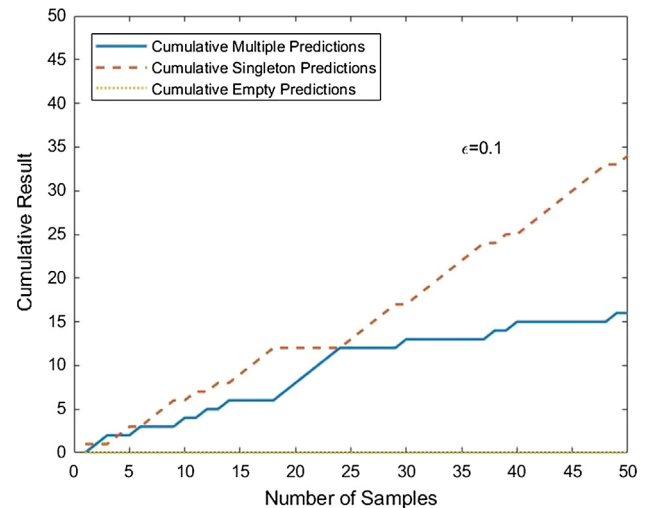
Sampel Indeks	BENAR Label	Dipaksa Ramalan	Kepercayaan diri	Kredibilitas	Sederhana Ramalan
1	0	0	0,9167	0,1667	0
25	0	0	0,9722	0,3333	0
37	0	0	0,9861	0,7083	0
59	1	1	0,9861	0,9306	1
64	1	1	0,9583	0,3194	1

3.2. Prediksi online dengan prediksi konformal

Menurut protokol yang diilustrasikan dalam Bagian2, CP-1NN dan CP-3NN digunakan untuk prediksi online. Sampel diacak ulang secara acak dan 20 sampel pertama digunakan sebagai set pelatihan awal. Setelah itu, setiap sampel baru dari kumpulan sampel yang telah dirombak menjalani prediksi.

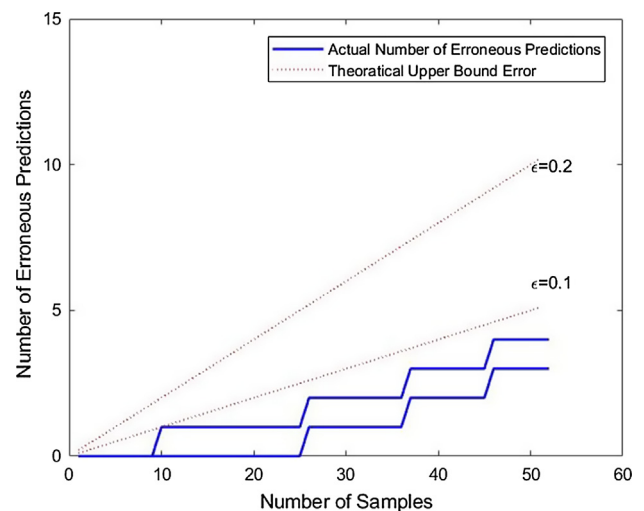
Hasil prediksi tunggal kumulatifDMenyanyi- nTH,banyak prediksiDBanyak- nTHdan prediksi kosongDEM- nTHdi bawah signifikansi-tingkat es 0,2 dan 0,1 diilustrasikan padaGambar 6 dan 7masing-masing. Awalnya, dengan sampel pelatihan yang terbatas, prediktor konformal cenderung menghasilkan banyak prediksi. Secara bertahap, seiring dengan bertambahnya observasi, model menjadi lebih kuat. Beberapa prediksi tidak lagi muncul saat prediksi tunggal berlaku, yang menunjukkan bahwa dengan lebih banyak sampel, sampel dapat diklasifikasikan ke dalam satu kelompok sambil menolak pilihan pasangan dengan keyakinan lebih tinggi. Kemungkinan alasannya adalah semakin banyak sampel yang memperluas ruang yang dicakup oleh jenis sampel tertentu dalam ruang sampel sehingga membantu meningkatkan pengetahuan prediktor dan kemampuan generalisasi.

Selain kecenderungan prediksi yang dibahas di atas, dengan memberikan bentuk prediksi seperti ganda, kosong, atau tunggal, prediksi konformal memberikan informasi kepada pengguna tentang sebaran sampel dalam ruang sampel. Misalnya, sampel ke-20 diberikan beberapa prediksi, yang menunjukkan bahwa sampel ini mungkin berada di dekat garis batas antara dua label. Sementara itu, jika salah satu sampel diberi prediksi kosong, berarti tidak satu pun dari kedua label tersebut memberikan prediksi yang dapat diandalkan untuk sampel tersebut, dan sampel tersebut mungkin berfungsi sebagai outlier yang tidak sesuai dengan kedua kelompok. Mengenai implikasi diagnosis kanker paru-paru, ketika muncul prediksi ganda atau prediksi kosong, dokter harus berpikir dua kali sebelum membuat diagnosis karena tidak jelas apakah prediksi tersebut akurat dan dapat diandalkan.

Gambar 6. Hasil prediksi konformal online dengan 1NN, $\epsilon=0.2$.Gambar 7. Hasil prediksi konformal online dengan 1NN, $\epsilon=0.1$.

Dua fitur spesifik umumnya digunakan untuk menggambarkan kinerja prediksi konformal online. Pertama, validitas prediksi konformal mengacu pada frekuensi kesalahan prediksi: jumlah total prediksi yang salah dibagi dengan jumlah sampel. Kecenderungan tingkat kesalahan pada tingkat signifikansi 0,2 dan 0,1 ditunjukkan padaGambar 8. Berdasarkan hasil validitas prediksi konformal online dapat diwujudkan bahwa dengan akumulasi observasi, tingkat kesalahan cenderung tidak melebihi batas atas yang ditetapkan oleh tingkat signifikansi.

Berdasarkan validitas yang tergambar terlihat dengan tingkat signifikansi yang lebih rendah-dan tingkat kepercayaan yang lebih tinggi 1 --, tingkat kesalahan dijamin lebih rendah dengan sampel yang cukup. Namun demikian, hal itu terjadi

Gambar 8. Kecenderungan kesalahan prediksi dengan CP-1NN $\epsilon=0.2$; 0.1TH.

belum tentu berarti semakin tinggi tingkat kepercayaan 1 --, semakin baik prediktor konformalnya, ketika efisiensi prediksi konformal dipertimbangkan.

Untuk memenuhi persyaratan reliabilitas, prediktor konformal cenderung menghasilkan beberapa prediksi (prediksi ganda dalam soal ini) untuk menghindari kesalahan klasifikasi sampel dan mematuhi tingkat kepercayaan. Hal ini menyebabkan prediktor menjadi kurang efisien. Oleh karena itu, selama metode pengukuran ketidaksesuaian tetap ada, pengguna perlu mencapai keseimbangan antara keyakinan dan efisiensi. Secara umum, ada dua kriteria utama yang menentukan efisiensi prediktor konformal[49]: persentase beberapa prediksi secara keseluruhan sampel yang diujiBanyak- n/n , dinotasikan sebagai kriteria M ('M' untuk 'Beberapa'), dan jumlah rata-rata label prediksi di wilayah prediksi dari beberapa prediksi, $W_{Mult} - n/n$ yang dilambangkan dengan kriteria E ('E' untuk 'Kelebihan'). Berdasarkan dua kriteria ini, hasilnya tercantum dalamTabel 4dan kecenderungan beberapa prediksi ditunjukkan pada Gambar. 9 dan 10. Berdasarkan hasil, karena ini adalah masalah klasifikasi biner, jumlah rata-rata beberapa prediksi adalah 2. Memilih kedua label menunjukkan efisiensi penurunan prediktor konformal, yang memerlukan analisis diagnostik lebih lanjut untuk sampel ini untuk menghindari kesalahan klasifikasi berdasarkan persyaratan keandalan saat ini. . Sementara itu, terbukti bahwa tingkat kepercayaan yang lebih tinggi 1 -- menghasilkan tingkat beberapa prediksi yang lebih tinggi. Oleh karena itu, pengguna perlu menetapkan tingkat signifikansi yang tepat untuk menyeimbangkan efisiensi dan keandalan.

Terlebih lagi, ketika perbandingan dilakukan antara CP-1NN dan CP-3NN, terlihat jelas dari hasil bahwa dalam kasus pengklasifikasian sampel kanker paru-paru, CP-3NN cenderung menghasilkan beberapa prediksi yang lebih sedikit, yang menunjukkan bahwa lebih efisien bila dibandingkan dengan CP-1NN.

3.3. Perbandingan dengan penelitian sebelumnya

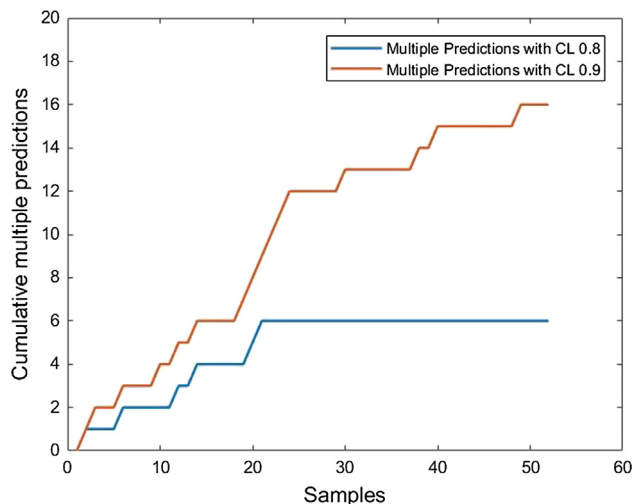
Penelitian sebelumnya telah dilakukan terkait skrining kanker paru dengan data hidung elektronik[50,51]. Misalnya, Rens van de Goor dkk. [50]menerapkan data hidung elektronik dan jaringan saraf tiruan (JST) dan berhasil membedakan pasien dengan kontrol dengan sensitivitas 88%, spesifisitas 86%, dan akurasi diagnostik 86%. Madara Tirzite dkk. [51]menggunakan analisis regresi logistik untuk mendeteksi kanker paru-paru dengan hidung elektronik dan memperoleh sensitivitas keseluruhan sebesar 95,8% untuk perokok dan 96,2% untuk bukan perokok.

Memang benar, dengan kumpulan data yang lebih besar, metrik prediksi pada penelitian-penelitian sebelumnya cenderung lebih baik. Jika dibandingkan dengan kedua penelitian ini, kontribusi spesifik dari makalah ini adalah sebagai berikut:

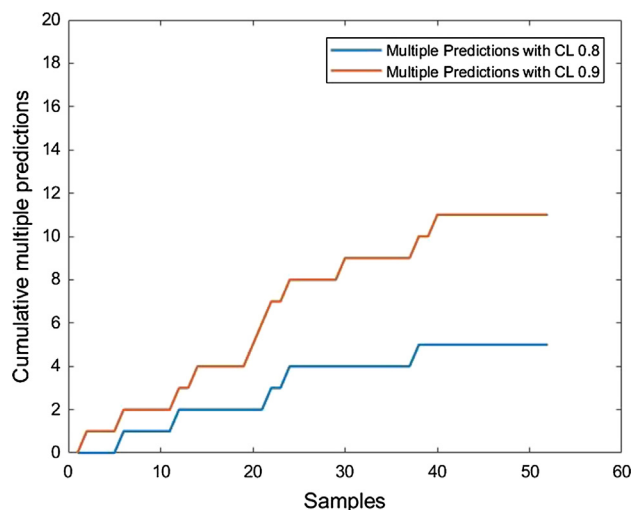
Pertama, sehubungan dengan kemampuan generalisasi dari satu kumpulan data ke kumpulan data lainnya, penelitian ini menawarkan informasi risiko dan ketidakpastian yang lebih baik untuk tugas klasifikasi di masa depan untuk kumpulan data yang tidak terlihat. Kumpulan data dan model cukup konsisten dengan penelitian sebelumnya, yang berarti para peneliti mendasarkan evaluasi mereka pada keseluruhan kelompok yang sama dalam eksperimen e-nose. Yang lebih penting adalah seberapa baik kinerja pengklasifikasi ketika menangani kumpulan data yang tidak homogen, seperti sistem e-nose dengan perbedaan produksi, suhu, kelembapan, dan kondisi penuaan sensor yang berbeda. Karena itu,

Tabel 4
Efisiensi prediksi konformal online.

Prediktor	BanyakN/N		W Multi-n/N	
	- $\frac{1}{4}0:2$	- $\frac{1}{4}0:1$	- $\frac{1}{4}0:2$	- $\frac{1}{4}0:1$
CP-1NN	0,1154	0,3077	2	2
CP-3NN	0,0962	0,2115	2	2



Gambar 9. Kecenderungan beberapa prediksi dengan CP-1NN.



Gambar 10. Kecenderungan beberapa prediksi dengan CP-3NN.

sensitivitasnya tinggi, spesifisitasnya bukan jaminan performa keseluruhan di berbagai kumpulan data. Karena tidak ada seorang pun yang dapat memeriksa semua kumpulan data hidung elektronik di dunia nyata, dalam hal ini, prediksi itu sendiri dan metrik yang terkait dengan prediksi dalam pengujian retrospektif tertentu mungkin tidak cukup tanpa mempertimbangkan keandalan, ketidakpastian informasi, dan keyakinan prediksi yang mungkin terjadi. diperluas untuk nilai yang lebih klinis. Penelitian saat ini lebih menekankan pada reliabilitas prediksi dengan CP-KNN. Hal ini bisa memberikan nilai lebih selain sekedar memberikan prediksi itu sendiri. Selain itu, landasan matematis dari keandalan Prediksi Konformal dibenarkan secara ketat oleh Vladimir Vovk. dan rekan kerja. Dalam penelitian ini, kami juga memvalidasi reliabilitas secara eksperimental secara online. Keyakinan dalam evaluasi keandalan dapat dipastikan, yang tidak dapat diinterpretasikan dalam ANN dan LR. ANN dan LR dapat bekerja dengan sangat baik dalam keadaan tertentu dengan beberapa sistem, kemampuan generalisasi, interpretasi keandalannya tidak diragukan lagi.

Kedua, kedua penelitian sebelumnya merupakan metode parametrik yang khas dan mengandalkan asumsi model tertentu. Untuk jaringan saraf tiruan, arsitektur, bobot, bias fungsi aktivasi yang dilatih dari set pelatihan mungkin tidak kompatibel dengan kasus dunia nyata jika ditransfer ke set data berbeda yang diperoleh dari

hidung elektronik yang berbeda. Hal ini dapat memerlukan lebih banyak waktu dalam proses pelatihan untuk setiap sistem hidung elektronik. Dalam regresi logistik, asumsi model linier dan estimasi kemungkinan maksimum dari parameter mungkin tidak sesuai untuk data hidung elektronik. Data dipengaruhi oleh gangguan sistem, penyimpangan sensor, dan dipengaruhi oleh kolinearitas, yang dapat menyebabkan pengklasifikasian menjadi sangat tidak kuat. Misalnya, kolinearitas antara fitur untuk satu sensor dan seluruh sensor dapat menyebabkan koefisien model regresi logistik bervariasi dalam rentang yang luas dan menjadi tidak stabil secara numerik. Oleh karena itu, ketika berhadapan dengan hidung elektronik, metode non-parametrik seperti CP-KNN mungkin merupakan pilihan yang lebih bijaksana.

4. Kesimpulan

Dalam karya ini, diperkenalkan penerapan baru prediksi konformal dalam prediksi kanker paru-paru dengan sistem hidung elektronik. Sampel udara pernapasan dari pasien kanker paru-paru dan kontrol dikumpulkan dan dianalisis dalam sistem e-nose. Setelah itu, data diolah dengan prediksi konformal baik dalam mode offline maupun online. Pengukuran ketidaksesuaian untuk prediksi konformal didasarkan pada 1NN dan 3NN dalam karya ini. Dalam mode offline, akurasi prediksi konformal berdasarkan 1NN (CP-1NN) dan 3NN (CP-3NN) masing-masing adalah 87,50% dan 83,33%, sedikit lebih baik dibandingkan hasil yang diperoleh dari prediktor sederhana 1NN dan 3NN. Selain hasil prediksi, prediktor konformal memungkinkan pengguna mengetahui keandalan prediksi individu dengan memberikan keyakinan dan kredibilitas setiap prediksi, yang penting dalam diagnosis kanker. Dalam mode online, validitas prediksi konformal diwujudkan dengan semakin banyaknya sampel, tingkat prediksi yang salah secara bertahap berada di bawah tingkat signifikansi yang ditetapkan oleh pengguna, yang berarti prediktor tidak hanya dapat secara efektif melakukan diagnosis kanker paru-paru secara online tetapi juga menjadi lebih akurat dan kuat. dengan sampel yang lebih banyak. Selain itu, potensi prediksi konformal untuk menunjukkan distribusi sampel telah dibahas. Sementara itu, pengguna perlu menyeimbangkan keyakinan dan efisiensi sambil mempertimbangkan tingkat keyakinan untuk prediksi konformal, karena tingkat keyakinan yang lebih tinggi biasanya menghasilkan beberapa prediksi yang lebih sering, yang merupakan simbol efisiensi yang lebih rendah. Perangkat lunak yang dikembangkan memberikan solusi analitis tambahan. Dengan perangkat keras hidung elektronik, data dapat diperoleh dan dianalisis dengan perangkat lunak di laboratorium dan klinik.

Di masa depan, metode pengukuran ketidaksesuaian yang berbeda dapat diterapkan dalam prediksi konformal untuk data yang dikumpulkan dari sistem hidung elektronik untuk meningkatkan validitas dan efisiensi karena proyek ini hanya mengusulkan kerangka kerja berdasarkan CP-KNN. Selain itu, pemilihan dan optimasi sensor buatan juga dapat dilakukan secara terawasi berdasarkan data untuk mengetahui sensor yang berkontribusi paling besar terhadap masalah klasifikasi. Terakhir, ketika menangani diagnosis dengan keandalan rendah, lebih banyak solusi seperti penggabungan data dengan metode analitik berbeda dan pencitraan diagnostik seperti GC-MS, CT untuk menangani kondisi dengan kredibilitas rendah juga dapat dipelajari untuk memberikan lebih banyak solusi. nilai diagnostik bagi pasien.

Pernyataan kontribusi kepenulisan CRediT

Xianghaozhan:Kursi Data, Investigasi, Perangkat Lunak, Penulisan - draf asli, Penulisan - review & editing.Zhan Wang:Kursi Data, Investigasi, Perangkat Lunak, Validasi, Penulisan - review & pengeditan. Meng Yang:Kursi Data, Investigasi, Perangkat Lunak, Validasi.Zhiyuan Luo:Pengawasan, Penulisan - review & editing. Anda Wang:Akuisisi pendanaan, Sumber Daya, Administrasi Proyek. Guang Li:Akuisisi pendanaan, Sumber Daya, Pengawasan, Administrasi Proyek.

Deklarasi Kepentingan Bersaing

Para penulis menyatakan bahwa mereka tidak mempunyai kepentingan finansial atau hubungan pribadi yang saling bersaing yang dapat mempengaruhi pekerjaan yang dilaporkan dalam makalah ini.

Ucapan Terima Kasih

Pekerjaan ini didukung oleh Natural Science Foundation of China (Hibah No. 61773342) dan Proyek Penelitian Otonom dari State Key Laboratory of Industrial Control Technology, China (Grant No. ICT1914).

Referensi

- [1] E. Office, Hari Kanker Sedunia: Lembar Fakta dan Promosi 3(1) (2017).
- [2] RL Siegel, KD Miller, A. Jemal, Statistik kanker, CA: A Cancer J. Dokter 67 (1) (2017) 5.
- [3] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, SY Ahn, dkk., Kematian global dan regional dari 235 penyebab kematian untuk 20 kelompok umur pada tahun 1990 dan 2010: analisis sistematis untuk studi beban penyakit global 2010, *Lancet* 380 (9859) (2012) 2095–2128.
- [4] P. Ong, D. Ost, Perubahan epidemiologi kanker paru-paru: Implikasi dalam diagnosis dan terapi.
- [5] LA Torre, F. Bray, RL Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Statistik kanker global, 2012, *Ca A Cancer J. Dokter* 65 (2) (2015) 87–108.
- [6] AJ Linz, JJ Picken, Deteksi dini kanker paru-paru: metode untuk meningkatkan tingkat kelangsungan hidup, *J. Am. Ahli osteopati. Asosiasi*. 79 (6) (1980) 364.
- [7] F. Taher, H. Al-Ahmad, N. Werghi, Deteksi dini kanker paru-paru berdasarkan analisis gambar berwarna dahak, dalam: *IEEE International Conference on Electronics*, 2014.
- [8] S. Peng, Q. Xu, XB Ling, X. Peng, W. Du, L. Chen, Klasifikasi molekuler jenis kanker dari data microarray menggunakan kombinasi algoritma genetika dan mesin vektor pendukung, *Febs Lett.* 555 (2) (2003) 358–362.
- [9] B. Yu, Y. Zhang, L. Zhao, Klasifikasi kanker dengan metode hybrid menggunakan data ekspresi gen microarray, *J. Comput. teori. ilmu nano*. 12 (10) (2015) 3194–3200.
- [10] J. Khan, JS Wei, M. Ringné, LH Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, CR Antonescu, C. Peterson, Klasifikasi dan prediksi diagnostik kanker menggunakan profil ekspresi gen dan jaringan saraf tiruan, *Nat. medis*. 7 (6) (2001) 673.
- [11] B. Yu, Q. Wang, X. Wang, S. Li, L. Lou, W. Qiu, Tentang ekstraksi gen informatif kanker dan penambahan data ekspresi gen, *J. Bionanosci.* 10 (4) (2016) 293–299 (7).
- [12] A. Bikov, Z. Lázár, I. Horvath, Masalah metodologis yang ada dalam penelitian hidung elektronik: seberapa jauh kita menggunakan instrumen ini dalam pengaturan klinis analisis napas? 9 (3) (2015) 034001.
- [13] FD Francesco, R. Fuoco, MG Trivella, A. Ceccarini, Analisis nafas: tren teknik dan aplikasi klinis, *Microchem. J.* 79 (1) (2005) 405–410.
- [14] B. Schmekel, F. Winquist, A. Vikström, Analisis sampel napas untuk kelangsungan hidup kanker paru-paru, *Anal. Chim. Undang-Undang* 840 (2014) 82–86.
- [15] D. Poli, M. Goldoni, M. Corradi, O. Acampa, P. Carbognani, E. Internullo, A. Casalini, A. Mutti, Penentuan aldehid pada hembusan nafas pasien kanker paru dengan cara on-fiber-derivatized spme-gc/ms, *J. Chromatogr. B Dubur. Teknologi. Bioma. Ilmu Kehidupan*. 878 (27) (2010) 2643–2651.
- [16] F. Wojciech, S. Andreas, F. Anna, A. Clemens, S. Jochen, M. Wolfram, A. Anton, T. Jakob, Td-gc-ms analisis metabolit volatil kanker paru-paru manusia dan sel normal in vitro, *Epidemiol Kanker. Biomarker* Sebelumnya. 19 (1) (2010) 182–195.
- [17] E. Gobbi, M. Falasconi, G. Zambotti, V. Sberveglieri, A. Pulvirenti, G. Sberveglieri, Diagnosis cepat enterobacteriaceae dalam sup sayuran dengan hidung elektronik berbasis sensor oksida logam, *Sens. Aktuator B Chem.* 207 (2015) 1104–1113.
- [18] C. Olguin, N. Laguarda-Miró, L. Pascual, E. García-Breijo, R. Martínez-Mañez, J. Soto, Hidung elektronik untuk mendeteksi sarin, soman dan tabun meniru dan mengganggu agen, *Aktuator Sensor B kimia*. 202 (10) (2014) 31–37.
- [19] S. Sankaran, LR Khot, S. Panigrahi, Biologi dan aplikasi sistem penginderaan penciuman: review, *Sens. Aktuator B Chem.* 171–172 (8) (2012) 1–17.
- [20] GC Green, ADC Chan, H. Dan, M. Lin, Menggunakan hidung elektronik berbasis sensor oksida logam (mos) untuk diskriminasi bakteri berdasarkan koloni individu dalam suspensi, *Sens. Aktuator B Chem.* 152 (1) (2011) 21–28.
- [21] S. De Vito, M. Piga, L. Martinotto, G. Di Francia Co, pemantauan polusi perkotaan no2 dan nox dengan hidung elektronik yang dikalibrasi di lapangan dengan regularisasi bayesian otomatis, *Sens. Aktuator B: Chem.* 143 (1) (2009) 182–191.
- [22] L. Zhang, F. Tian, H. Nie, L. Dang, G. Li, Q. Ye, C. Kadri, Klasifikasi beberapa kontaminan udara dalam ruangan dengan hidung elektronik dan mesin vektor pendukung hibrida, *Sens. Aktuator B: kimia*. 174 (2012) 114–125.
- [23] R. Munoz, EC Sivret, G. Parcsi, R. Lebrero, X. Wang, IM Suffet, RM Stuetz, Teknik pemantauan untuk penilaian pengurangan bau, *Water Res.* 44 (18) (2010) 5129–5149.

- [24] T. Saidi, O. Zaim, M. Moufid, N. El Bari, R. Ionescu, B. Bouchikhi, Analisis pernafasan menggunakan kromatografi hidung dan gas elektronik – spektrometri massa untuk diagnosis non-invasif penyakit ginjal kronis, diabetes mellitus dan subyek sehat, *Sens. Aktuator B: Kimia*. 257 (2018) 178–188.
- [25] P. Montuschi, N. Mores, A. Trové, C. Mondino, P.J Barnes, Hidung elektronik dalam pengobatan pernafasan, *Respirasi* 85 (1) (2013) 72–84.
- [26] A.K Pavlou, N. Magan, C. McNulty, J.M Jones, D. Sharp, J. Brown, A.P Turner, Penggunaan sistem hidung elektronik untuk diagnosis infeksi saluran kemih, *Biosens. Bioelektron.* 17 (10) (2002) 893–899.
- [27] V.S Kodogiannis, J.N Lygouras, A. Tarczynski, H.S Chowdrey, Sistem diskriminasi bau buatan menggunakan hidung elektronik dan jaringan saraf untuk identifikasi infeksi saluran kemih, *IEEE Trans. Memberitahukan. Teknologi. Bioma.* 12 (6) (2008) 707–713.
- [28] V.V Musatov, V. Sysoev, M. Sommer, I. Kiselev, Penilaian kesegaran daging dengan sensor oksida logam microarray hidung elektronik: pendekatan praktis, *Sens. Aktuator B: Chem.* 144 (1) (2010) 99–103.
- [29] J. Ragazzo-Sanchez, P. Chaliel, D. Chevalier-Lucia, M. Calderon-Santoyo, C. Ghommidh, Deteksi rasa tidak enak dalam minuman beralkohol dengan hidung elektronik yang digabungkan dengan gc, *Sens. Aktuator B: Chem.* 140 (1) (2009) 29–34.
- [30] A. Loutfi, S. Coradeschi, G.K Mani, P. Shankar, J.B.B Rayappan, Hidung elektronik untuk kualitas makanan: review, *J. Food Eng.* 144 (2015) 103–111.
- [31] M. Tohidi, M. Ghasemi-Varnamkhasti, V. Ghafarinia, S.S Mohtasebi, M. Bonyadian, Identifikasi sejumlah kecil bubuk deterjen dalam susu mentah menggunakan sistem penciuman buatan berbiaya rendah yang disesuaikan: metode baru, *Pengukuran* 124 (2018) 120–129.
- [32] S. Faal, M. Loghavi, S. Kamgar, Sifat fisikokimia madu ziziphus Iran dan pendekatan yang muncul untuk memprediksinya menggunakan hidung elektronik, *Pengukuran* 148 (2019) 106936.
- [33] M. Ezhilan, N. Nesakumar, K.J Babu, C. Srinandan, J.B.B Rayappan, Penilaian kesegaran brokoli menggunakan electronic nose, *Pengukuran* 145 (2019) 735–743.
- [34] A. Sanaeifar, S.S Mohtasebi, M. Ghasemi-Varnamkhasti, H. Ahmadi, Penerapan hidung elektronik berbasis mos untuk prediksi sifat kualitas pisang, *Pengukuran* 82 (2016) 105–114.
- [35] A. Gammerman, V. Vovk, Prediksi lindung nilai dalam pembelajaran mesin pada kuliah jurnal komputer kedua, *Comput. J.* 10 (2) (2007) 151–163.
- [36] I. Nourredinov, D. Devetyarov, V. Vovk, B. Burford, S. Camuzeaux, A. Gentry-Maharaj, A. Tiss, C. Smith, Z. Luo, A. Chervonenkis, Prediksi multiprobabilistik dalam diagnosis medis awal, *Ann . Matematika. Artifi. Intel.* 74 (1–2) (2015) 1–20.
- [37] C. Zhou, I. Nourredinov, Z. Luo, D. Adamskiy, L. Randell, N. Coldham, A. Gammerman, Perbandingan Mesin Venn dengan Metode Platt dalam Output Probabilistik, Springer, Berlin, Heidelberg, 2011.
- [38] V. Vovk, Validitas bersyarat dari prediktor konformal induktif, *Mach. Mempelajari.* 92 (2–3) (2013) 349–376.
- [39] V. Vovk, A. Gammerman, G. Shafer, Pembelajaran algoritmik di dunia acak, 2005 xvi.
- [40] Z. Wang, X. Sun, J. Miao, Y. Wang, Z. Luo, G. Li, Prediksi konformal berdasarkan k-tetangga terdekat untuk diskriminasi ginseng dengan hidung elektronik buatan sendiri, *Sensor* 17 (8) (2017) 1869.
- [41] X. Sun, L. Liu, Z. Wang, J. Miao, Y. Wang, Z. Luo, G. Li, Pembelajaran ansambel multiklasifikasi yang dioptimalkan untuk identifikasi ginseng berdasarkan hidung elektronik, *Sens. Actuators A Phys.* 266.
- [42] J. Miao, Z. Luo, Y. Wang, G. Li, Perbandingan dan fusi data hidung elektronik dan spektroskopi reflektansi inframerah dekat untuk diskriminasi ginseng, *Anal. Metode* 8 (6) (2016) 1265–1273.
- [43] Y. Wang, J. Miao, X. Lyu, L. Liu, Z. Luo, G. Li, Prediksi probabilistik yang valid untuk ginseng dengan mesin venn menggunakan hidung elektronik, *Sensor* 16 (7) (2016) 1088.
- [44] Z. Haddi, M. Boughrini, S. Ihlou, A. Amari, Klasifikasi geografis minyak zaitun murni dengan menggabungkan hidung dan lidah elektronik, dalam: *Sensors*, 2012, hlm. 1–4.
- [45] K. Timsorn, C. Wongchoosuk, P. Wattuya, S. Promdaen, S. Sittichat, Diskriminasi kesegaran ayam menggunakan hidung elektronik dikombinasikan dengan pca dan ann, dalam: *Konferensi Internasional Teknik Elektro/elektronik, Komputer, Telekomunikasi dan Teknologi Informasi*, 2014, hlm.1–4.
- [46] X. Zhan, X. Guan, R. Wu, Z. Wang, Y. Wang, G. Li, Diskriminasi obat herbal alternatif dari kategori berbeda dengan hidung elektronik, *Sensor* 18 (9) (2018) 2936.
- [47] X. Zhan, X. Guan, R. Wu, Z. Wang, Y. Wang, Z. Luo, G. Li, Prediksi konformal online untuk mengklasifikasikan berbagai jenis obat herbal dengan hidung elektronik.
- [48] X. Zhan, X. Guan, R. Wu, Z. Wang, Y. Wang, G. Li, Rekayasa fitur dalam diskriminasi obat-obatan herbal dari asal geografis yang berbeda dengan hidung elektronik, dalam: *Konferensi Internasional IEEE ke-7 2019 tentang Bioinformatika dan Biologi Komputasi (ICBCB)*, IEEE, 2019, hlm.56–62.
- [49] V. Vovk, V. Fedorova, I. Nourredinov, A. Gammerman, Kriteria efisiensi untuk prediksi konformal (2016) 23–39.
- [50] R.van de Goor, M.van Hooren, A.-M. Dingemans, B. Kremer, K. Kross, Pelatihan dan validasi hidung elektronik portabel untuk skrining kanker paru-paru, *J. Thoracic Oncol.* 13 (5) (2018) 676–681.
- [51] M. Tirzite, M. Bukovskis, G. Strazda, N. Jurka, I. Taivans, Deteksi kanker paru-paru dengan analisis regresi hidung dan logistik elektronik, *J. Breath Res.* 13 (1) (2018) 016006.