Eur. Phys. J. C (2022) 82:283
https://doi.org/10.1140/epjc/s10052-022-10180-9

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

B-flavor tagging at Belle II

F. Abudinén[25], N. Akopov[64], A. Aloisio[19,43], V. Babu[5], Sw. Banerjee[56], M. Bauer[26], J. V. Bennett[59], F. U. Bernlochner[51], M. Bessner[54], S. Bettarini[22,46], T. Bilka[4], S. Bilokin[29], D. Biswas[56], D. Bodrov[12,28], J. Borah[14], M. Bračko[39,57], P. Branchini[23], A. Budano[23], M. Campajola[43,19], G. Casarosa[22,46], C. Cecchi[21,45], R. Cheaib[5], V. Chekelian[31], C. Chen[30], Y. Q. Chen[13], H. -E. Cho[9], S. Cunliffe[5], G. De Nardo[19,43], G. De Pietro[23], R. de Sangro[18], S. Dey[41], A. Di Canto[2], F. Di Capua[43,19], T. V. Dong[7], G. Dujany[50], P. Ecker[26], M. Eliachevitch[51], T. Ferber[26], F. Forti[22,46], E. Ganiev[25,49], A. Gaz[20,44], M. Gelb[26], J. Gemmler[26], R. Godang[60], P. Goldenzweig[26], E. Graziani[23], K. Hara[8,10], A. Hershenhorn[52], T. Higuchi[61], E. C. Hill[52], M. Hohmann[58], T. Humair[31], G. Inguglia[16], H. Junkerkalefeld[51], R. Karl[5], Y. Kato[32,33], T. Keck[26], C. Kiesling[31], C. -H. Kim[9], S. Kohani[54], I. Komarov[5], T. M. G. Kraetzschmar[31], P. Križan[39,55], J. F. Krohn[58], T. Kuhr[29], J. Kumar[3], K. Kumara[63], S. Kurz[5], S. Lacaprara[20], C. La Licata[61], M. Laurenza[23,47], K. Lautenbach[1], S. C. Lee[27], K. Lieret[29], L. Li Gioi[31], Q. Y. Liu[5], S. Longo[5], M. Maggiora[24,48], E. Manoni[21], C. Marinas[17], A. Martini[5], F. Meier[6], M. Merola[19,43], F. Metzner[26], M. Milesi[58], K. Miyabayashi[35], G. B. Mohanty[40], F. Mueller[31], C. Murphy[61], E. R. Oxford[3], S. -H. Park[10], A. Passeri[23], F. Pham[58], L. E. Piilonen[62], S. Pokharel[59], M. T. Prim[51], C. Pulvermacher[26], P. Rados[5], M. Ritter[29], A. Rostomyan[5], S. Sandilya[15], L. Santelj[55,39], Y. Sato[42], A. J. Schwartz[53], M. E. Sevior[58], A. Soffer[41], S. Spataro[24,48], R. Stroili[20,44], W. Sutcliffe[51], D. Tagnani[23], M. Takizawa[37,11,38], U. Tamponi[24], F. Tenchini[22,46], E. Torassa[20], P. Urquijo[58], L. Vitale[25,49], Y. Yusa[36], L. Zani[1], Q. D. Zhou[32,33,34], R. Žlebčík[4], A. Zupanc[39]

[1] Aix Marseille Université, CNRS/IN2P3, CPPM, 13288 Marseille, France
[2] Brookhaven National Laboratory, Upton, New York 11973, USA
[3] Carnegie Mellon University, Pittsburgh, PA 15213, USA
[4] Faculty of Mathematics and Physics, Charles University, 121 16 Prague, Czech Republic
[5] Deutsches Elektronen-Synchrotron, 22607 Hamburg, Germany
[6] Duke University, Durham, NC 27708, USA
[7] Institute of Theoretical and Applied Research (ITAR), Duy Tan University, Hanoi 100000, Vietnam
[8] The Graduate University for Advanced Studies (SOKENDAI), Hayama 240-0193, Japan
[9] Department of Physics and Institute of Natural Sciences, Hanyang University, Seoul 04763, South Korea
[10] High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan
[11] J-PARC Branch, KEK Theory Center, High Energy Accelerator Research Organization (KEK), Tsukuba, 305-0801, Japan
[12] Higher School of Economics (HSE), Moscow, 101000, Russian Federation
[13] Hunan Normal University, Changsha 410081, China
[14] Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India
[15] Indian Institute of Technology Hyderabad, Hyderabad, Telangana 502285, India
[16] Institute of High Energy Physics, 1050 Vienna, Austria
[17] Instituto de Fisica Corpuscular, 46980 Paterna, Spain
[18] INFN Laboratori Nazionali di Frascati, 00044 Frascati, Italy
[19] INFN Sezione di Napoli, 80126 Naples, Italy
[20] INFN Sezione di Padova, 35131 Padua, Italy
[21] INFN Sezione di Perugia, 06123 Perugia, Italy
[22] INFN Sezione di Pisa, 56127 Pisa, Italy
[23] INFN Sezione di Roma Tre, 00146 Rome, Italy
[24] INFN Sezione di Torino, 10125 Turin, Italy
[25] INFN Sezione di Trieste, 34127 Trieste, Italy
[26] Institut für Experimentelle Teilchenphysik, Karlsruher Institut für Technologie, 76131 Karlsruhe, Germany
[27] Kyungpook National University, Daegu 41566, South Korea

[28] P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow 119991, Russian Federation
[29] Ludwig Maximilians University, 80539 Munich, Germany
[30] Iowa State University, Ames, IA 50011, USA
[31] Max-Planck-Institut für Physik, 80805 Munich, Germany
[32] Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan
[33] Kobayashi-Maskawa Institute, Nagoya University, Nagoya 464-8602, Japan
[34] Institute for Advanced Research, Nagoya University, Nagoya, 464-8602, Japan
[35] Nara Women's University, Nara, 630-8506, Japan
[36] Niigata University, Niigata, 950-2181, Japan
[37] Meson Science Laboratory, Cluster for Pioneering Research, RIKEN, Saitama, 351-0198, Japan
[38] Showa Pharmaceutical University, Tokyo, 194-8543, Japan
[39] J. Stefan Institute, 1000 Ljubljana, Slovenia
[40] Tata Institute of Fundamental Research, Mumbai 400005, India
[41] School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
[42] Department of Physics, Tohoku University, Sendai 980-8578, Japan
[43] Dipartimento di Scienze Fisiche, Università di Napoli Federico II, 80126 Naples, Italy
[44] Dipartimento di Fisica e Astronomia, Università di Padova, 35131 Padua, Italy
[45] Dipartimento di Fisica, Università di Perugia, 06123 Perugia, Italy
[46] Dipartimento di Fisica, Università di Pisa, 56127 Pisa, Italy
[47] Dipartimento di Matematica e Fisica, Università di Roma Tre, 00146 Rome, Italy
[48] Dipartimento di Fisica, Università di Torino, 10125 Turin, Italy
[49] Dipartimento di Fisica, Università di Trieste, 34127 Trieste, Italy
[50] Université de Strasbourg, CNRS, IPHC, UMR 7178, 67037 Strasbourg, France
[51] University of Bonn, 53115 Bonn, Germany
[52] University of British Columbia, Vancouver, BC V6T 1Z1, Canada
[53] University of Cincinnati, Cincinnati, OH 45221, USA
[54] University of Hawaii, Honolulu, HI 96822, USA
[55] Faculty of Mathematics and Physics, University of Ljubljana, 1000 Ljubljana, Slovenia
[56] University of Louisville, Louisville, KY 40292, USA
[57] University of Maribor, 2000 Maribor, Slovenia
[58] School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[59] University of Mississippi, University, MS 38677, USA
[60] University of South Alabama, Mobile, AL 36688, USA
[61] Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo, Kashiwa 277-8583, Japan
[62] Virginia Polytechnic Institute and State University, Blacksburg, VI 24061, USA
[63] Wayne State University, Detroit, MI 48202, USA
[64] Alikhanyan National Science Laboratory, 0036 Yerevan, Armenia

Received: 14 September 2021 / Accepted: 1 March 2022 / Published online: 1 April 2022

Abstract We report on new flavor tagging algorithms developed to determine the quark-flavor content of bottom (B) mesons at Belle II. The algorithms provide essential inputs for measurements of quark-flavor mixing and charge-parity violation. We validate and evaluate the performance of the algorithms using hadronic B decays with flavor-specific final states reconstructed in a data set corresponding to an integrated luminosity of $62.8\,\text{fb}^{-1}$, collected at the $\Upsilon(4S)$ resonance with the Belle II detector at the SuperKEKB collider. We measure the total effective tagging efficiency to be
$\varepsilon_{\text{eff}} = \big(30.0 \pm 1.2(\text{stat}) \pm 0.4(\text{syst})\big)\%$
for a category-based algorithm and
$\varepsilon_{\text{eff}} = \big(28.8 \pm 1.2(\text{stat}) \pm 0.4(\text{syst})\big)\%$
for a deep-learning-based algorithm.

1 Introduction

Determining the quark-flavor content of heavy-flavored hadrons is essential in many measurements of quark-flavor mixing and CP violation. A keystone of the Belle II physics program is the study of $B^0 - \bar{B}^0$ mixing and CP violation in decays of neutral B mesons [1–21]. The study of these processes is key to constrain the Cabibbo–Kobayashi–Maskawa (CKM) angles ϕ_1/β and ϕ_2/α [22–27], as well as

to study flavor anomalies that could ultimately reveal possible deviations from the Standard Model expectations [28–30].

At Belle II, B mesons are produced in $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance, which decays almost half of the time into a pair of neutral $B\bar{B}$ mesons. Most measurements of CP violation and $B^0 - \bar{B}^0$ mixing require the full reconstruction of a signal B meson (signal side), and to determine the quark-flavor content of the accompanying B meson (tag side) at the time of its decay, a task referred to as flavor tagging.

Flavor tagging is possible because many decay modes of neutral B mesons provide flavor signatures through flavor-specific final states. Flavor signatures are characteristics of the B-decay products that are correlated with the flavor of the neutral B meson, which is the charge sign of the b quark or antiquark that it contains. For example, in semileptonic decays such as $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ (charge-conjugate processes are implied everywhere in this paper), a negatively charged lepton tags unambiguously a $\overline{B}^0$, which contains a negatively charged b, while a positively charged lepton tags a B^0, which contains a positively charged $\bar{b}$.

To determine the quark-flavor of B mesons, Belle II exploits the information associated with the B-decay products using multivariate machine-learning algorithms. We develop two algorithms. The first one is a category-based flavor tagger [31], which is inspired by previous flavor taggers developed by the Belle and the BaBar collaborations [32,33]. The category-based flavor tagger first identifies B^0-decay products and then combines all information to determine the B^0 flavor. The second algorithm is a deep-learning neural network (DNN) flavor tagger [34], that determines the B^0 flavor in a single step without pre-identifying B^0-decay products.

In the following, we focus on the description of the algorithms and their training procedure. We evaluate the performance of the algorithms by measuring the $B^0 - \bar{B}^0$ mixing probability from the signal yield integrated in decay-time (time-integrated analysis).

To evaluate the performance, we reconstruct signal B decays with final states that allow us to unambiguously identify the flavor of the signal side and determine the flavor of the tag side using the flavor taggers. We reconstruct signal B decays into hadronic final states with branching fractions of 10^{-5} or greater to obtain a sufficiently large signal sample in the used data set. We evaluate the tagging performance on neutral B mesons and, as a cross check, on charged B mesons. The training of the flavor taggers, the signal B reconstruction procedure and the event selection criteria are developed and optimized using Monte Carlo (MC) simulation before application to experimental data.

The remainder of this paper is organized as follows. Section 2 describes the Belle II detector, followed by a description of the data sets and analysis framework in Sect. 3. The category-based flavor tagger is described in Sect. 4 and the DNN tagger in Sect. 5. The training of the flavor taggers is detailed in Sect. 6. We describe the reconstruction of the calibration samples in Sect. 7 and the determination of efficiencies and wrong-tag fractions in Sect. 8. We then compare the performance of the flavor taggers in data and in simulation in Sect. 9 and present the results of the calibration in Sects. 10 and 11. A comparison with the Belle algorithm is provided in Sect. 12, followed by a summary of the paper in Sect. 13.

2 The Belle II detector

Belle II is a particle-physics spectrometer with a solid-angle acceptance of almost 4π [1,35]. It is designed to reconstruct the products of electron–positron collisions produced by the SuperKEKB asymmetric-energy collider [36], located at the KEK laboratory in Tsukuba, Japan. Belle II comprises several subdetectors arranged around the interaction point in a cylindrical geometry. The innermost subdetector is the vertex detector, which uses position-sensitive silicon layers to sample the trajectories of charged particles (tracks) in the vicinity of the interaction region to determine the decay vertex of their long-lived parent particles. The vertex detector includes two inner layers of silicon pixel sensors (PXD) and four outer layers of double-sided silicon microstrip sensors (SVD). The second pixel layer is currently incomplete and covers only one sixth of azimuthal angle. Charged-particle momenta and charges are measured by a large-radius, helium-ethane, small-cell central drift chamber (CDC), which also offers charged-particle-identification information via the measurement of particles' energy-loss $\mathrm{d}E/\mathrm{d}x$ by specific ionization. A Cherenkov-light angle and time-of-propagation detector (TOP) surrounding the chamber provides charged-particle identification in the central detector volume, supplemented by proximity-focusing, aerogel, ring-imaging Cherenkov detectors (ARICH) in the forward regions. A CsI(Tl)-crystal electromagnetic calorimeter (ECL) enables energy measurements of electrons and photons. A solenoid surrounding the calorimeter generates a uniform axial 1.5 T magnetic field filling its inner volume. Layers of plastic scintillator and resistive-plate chambers (KLM), interspersed between the magnetic flux-return iron plates, enable identification of K^0_L mesons and muons. We employ all subdetectors in this work.

3 Framework and data

Both flavor taggers are part of the Belle II analysis software framework [37], which is used to process all data. We train the flavor taggers using a sample of 20 million

signal-only MC events [38,39], where the signal B meson decays to the invisible final state $B^0 \to \nu_\tau \bar{\nu}_\tau$ and the tag-side B meson decays to any possible final state according to known branching fractions [40]. To perform different tests after training, we use similar signal-only MC samples where the signal B meson decays to benchmark decay modes such as $B^0 \to K^+\pi^-$, $B^0 \to J/\psi(\to \mu^+\mu^-)\, K^0_S(\to \pi^+\pi^-)$, and $B^0 \to \eta'(\to \pi^+\pi^-\eta(\to \pi^+\pi^-\pi^0))\, K^0_S(\to \pi^+\pi^-)$.

We evaluate the performance of the flavor taggers using *generic MC simulation*. The generic MC simulation consists of samples that include $e^+e^- \to B^0\bar{B}^0$, B^+B^-, $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, and $s\bar{s}$ processes [38–41] in proportions representing their different production cross sections and correspond to an integrated luminosity of 700 fb^{-1}, about eleven times the data sample used in the measurement. We use these samples to optimize the event selection and to compare the flavor distributions and fit results obtained from the experimental data with the expectations.

Signal-only and generic MC simulation include the effect of simulated beam-induced backgrounds [1,42], caused by the Touschek effect (scattering and loss of beam particles), by beam-gas scattering and by synchrotron radiation, as well as simulated luminosity-dependent backgrounds, caused by Bhabha scattering [43,44] and by two-photon quantum electrodynamical processes [45].

As for experimental data, we use all good-quality data collected at the $\Upsilon(4S)$ resonance between March 11$^{\text{th}}$, 2019 and July 1$^{\text{st}}$, 2020; this sample corresponds to an integrated luminosity of 62.8 fb^{-1} [46]. To reduce the data sample size to a manageable level, all events are required to meet loose data-skim selection criteria, based on total energy and charged-particle multiplicity in the event. Almost 100% of the $B\bar{B}$ events meet the data-skim selection criteria.

4 The category-based flavor tagger

The Belle II flavor taggers are multivariate algorithms that receive as input kinematic, track-hit, and charged-particle identification (PID) information about the particles on the tag side and provide as output the product $q \cdot r$, where q is the flavor of the tag-side B meson, and r the dilution factor. By definition, the dilution factor r is equal to $1 - 2w$, where w corresponds to the fraction of wrongly tagged events. A dilution factor $r = 0$ indicates a fully diluted flavor (no possible distinction between B^0 and $\bar{B}^0$), whereas a dilution factor $r = 1$ indicates a perfectly tagged flavor. By convention, $q = +1$ corresponds to a tag-side B^0, and $q = -1$ corresponds to a tag-side $\bar{B}^0$.

The new category-based algorithm relies on flavor-specific decay modes. Each decay mode has a particular decay topology and provides a flavor-specific signature. Similar or complementary decay modes are combined to obtain additional flavor signatures. The different flavor signatures are sorted into thirteen *tagging categories*, which are described in detail in Sect. 4.1. Table 1 shows an overview of all thirteen categories together with the underlying decay modes and the respective flavor-specific final state particles, which we call *target particles*.

We identify the target particles among all available particle candidates on the tag side using discriminating input variables. Some input variables require information from all reconstructed tracks (charged candidates) [47] and all neutral clusters (neutral candidates) on the tag side. Neutral clusters are clusters in the electromagnetic calorimeter (reconstructed photons) and in the KLM detector (reconstructed K^0_L particles) that are not associated with a reconstructed track. Summing the input variables for all categories yields a total of 186 inputs in the current configuration of the algorithm (see Sect. 4.1). Some variables are used multiple times for the same candidates in different categories. To save computing time, each variable is calculated only once for each candidate.

We adopted the useful concept of tagging categories from the previous Belle and BaBar flavor taggers [32,33]. However, the new Belle II category-based flavor tagger includes more categories and more input variables than previously implemented algorithms.

4.1 Categories and input variables

In the following, we describe the flavor signatures and the input variables. Table 2 shows an overview of the input variables for each category. Except for the Maximum-p^* category, PID variables are used for all categories. The PID variables correspond to PID likelihoods $\mathcal{L}$ [1], which can be either combined likelihoods considering 6 possible long-lived charged-particle hypotheses (e, μ, K, π, p, and deuteron), or binary likelihoods considering only two of the hypotheses. The PID likelihoods can be calculated using all sub-detectors providing particle identification, or single ones (TOP, ARICH, ECL, KLM, or dE/dx from CDC). For example, $\mathcal{L}^{\mathrm{d}E/\mathrm{d}x}_{\pi/e}$ stands for binary π/e PID using only CDC information.

Electron, muon, and kinetic lepton These categories exploit the signatures provided by primary leptons from B decays occurring via transitions $b \to c\, \ell^- \bar{\nu}_\ell$, or $b \to u\, \ell^- \bar{\nu}_\ell$, where ℓ corresponds to an electron, muon or both depending on the category. Useful variables to identify primary leptons are the momentum p, the transverse momentum p_t, and the cosine of the polar angle $\cos\theta$, which can be calculated in the lab frame, or in the $\Upsilon(4S)$ frame (denoted with a * superscript). We consider the following variables calculated only in the $\Upsilon(4S)$ frame:

Table 1 Tagging categories and their targets (left) with examples of the considered decay modes (right). The target particles for each category are shown using the same colors on the left and on the right. Here, p^* stands for momentum in the center-of-mass frame, $\ell^\pm$ for charged leptons (μ^- or e^-) and X for other possible particles in the decays

Categories	Targets for $\overline{B}^0$
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
Kinetic Lepton	ℓ^-
Intermediate Kinetic Lepton	ℓ^+
Kaon	K^-
Kaon-Pion	K^-, π^+
Slow Pion	π^+
Maximum p^*	ℓ^-, π^-
Fast-Slow-Correlated (FSC)	ℓ^-, π^+
Fast Hadron	π^-, K^-
Lambda	Λ

Underlying decay modes

$\overline{B}^0 \to D^{*+}\ \bar{\nu}_\ell\ \ell^-$
$\quad\hookrightarrow D^0\ \pi^+$
$\quad\quad\hookrightarrow X\ K^-$

$\overline{B}^0 \to D^+\ \pi^-\ (K^-)$
$\quad\hookrightarrow \overline{K}^0\ \nu_\ell\ \ell^+$

$\overline{B}^0 \to \Lambda_c^+\ X^-$
$\quad\hookrightarrow \Lambda\ \pi^+$
$\quad\quad\hookrightarrow p\ \pi^-$

Table 2 Discriminating input variables for each category and for the DNN flavor tagger. For some of the categories the p-value of the track fit is taken into account. For the Lambda category, the p-value of the reconstructed Λ-decay vertex is used. All variables are calculated for every particle candidate

Categories	Discriminating input variables
Electron and Int. Electron	$\mathcal{L}_e$,$\mathcal{L}_e^{\text{TOP}}$, $\mathcal{L}_e^{\text{ARICH}}$, $\mathcal{L}_e^{\text{ECL}}$, p^*, p_t^*, p, p_t, $\cos\theta$, M^2_{rec}, E^W_{90}, p^*_{miss}, $\cos\theta^*_{\text{miss}}$, $\lvert\cos\theta^*_T\rvert$, p-val.
Muon and Int. Muon	$\mathcal{L}_\mu$,$\mathcal{L}_\mu^{\text{TOP}}$, $\mathcal{L}_\mu^{\text{ARICH}}$, $\mathcal{L}_\mu^{\text{KLM}}$, p^*, p_t^*, p, p_t, $\cos\theta$, M^2_{rec}, E^W_{90}, p^*_{miss}, $\cos\theta^*_{\text{miss}}$, $\lvert\cos\theta^*_T\rvert$, p-val.(only for Int.)
Kin. Lepton and Int. Kin. Lep.	$\mathcal{L}_e$, $\mathcal{L}_e^{\text{TOP}}$, $\mathcal{L}_e^{\text{ARICH}}$, $\mathcal{L}_e^{\text{ECL}}$, $\mathcal{L}_\mu$, $\mathcal{L}_\mu^{\text{TOP}}$, $\mathcal{L}_\mu^{\text{ARICH}}$, $\mathcal{L}_\mu^{\text{KLM}}$, p^*, p_t^*, p, p_t, $\cos\theta$, M^2_{rec}, E^W_{90}, p^*_{miss}, $\cos\theta^*_{\text{miss}}$, $\lvert\cos\theta^*_T\rvert$, p-val.(only for Int.)
Kaon	$\mathcal{L}_K$, $\mathcal{L}_K^{dE/dx}$, $\mathcal{L}_K^{\text{TOP}}$, $\mathcal{L}_K^{\text{ARICH}}$, p^*, p_t^*, p, p_t, $\cos\theta$, $n_{K^0_S}$, $\sum p_t^2$, M^2_{rec}, E^W_{90}, p^*_{miss}, $\cos\theta^*_{\text{miss}}$, $\lvert\cos\theta^*_T\rvert$, p-val.
Slow Pion and Fast Hadron	$\mathcal{L}_\pi$,$\mathcal{L}_\pi^{\text{TOP}}$, $\mathcal{L}_\pi^{\text{ARICH}}$, $\mathcal{L}_{\pi/e}^{dE/dx}$, $\mathcal{L}_e$, $\mathcal{L}_K$, $\mathcal{L}_K^{dE/dx}$, $\mathcal{L}_K^{\text{TOP}}$, $\mathcal{L}_K^{\text{ARICH}}$, p^*, p_t^*, p, p_t,$\cos\theta$,$n_{K^0_S}$, $\sum p_t^2$, M^2_{rec},E^W_{90}, p^*_{miss}, $\cos\theta^*_{\text{miss}}$, $\lvert\cos\theta^*_T\rvert$
Kaon-Pion	$\mathcal{L}_K$, y_{Kaon}, y_{SlowPion}, $\cos\theta^*_{K\pi}$, $q_K \cdot q_\pi$
Maximum p^*	p^*, p_t^*, p, p_t, $\lvert\cos\theta^*_T\rvert$
FSC	$\mathcal{L}_K$, p^*_{Slow}, p^*_{Fast}, $\lvert\cos\theta^*_{T,\text{Slow}}\rvert$, $\lvert\cos\theta^*_{T,\text{Fast}}\rvert$, $\cos\theta^*_{\text{SlowFast}}$, $q_{\text{Slow}} \cdot q_{\text{Fast}}$
Lambda	$\mathcal{L}_p$, $\mathcal{L}_\pi$, p^*_Λ, p_Λ, p^*_π, p_p, p^*_π,p_π,q_Λ,M_Λ, $n_{K^0_S}$,$\cos\theta_{\boldsymbol{x}_\Lambda,\boldsymbol{p}_\Lambda}$, σ^{zz}_Λ, p-val.
DNN	$\mathcal{L}_e$, $\mathcal{L}_\mu$, $\mathcal{L}_K$, $\mathcal{L}_\pi$, $\mathcal{L}_p$, p^*, $\cos(\theta^*)$, ϕ^*, N_{PXD}, N_{SVD}

- $M^2_{\text{rec}} = m_X^2 = g_{\mu,\nu} p_X^{*\mu} p_X^{*\nu}$, the squared invariant mass of the recoiling system X whose four-momentum is defined by

$$p_X^{*\mu} = \sum_{i \neq \ell} p_i^{*\mu},$$

  where the index i goes over all charged and neutral candidates on the tag side and ℓ corresponds to the index of the lepton candidate.
- $p^*_{\text{miss}} = |\boldsymbol{p}^*_{\text{miss}}| = |\boldsymbol{p}^*_{B^0_{\text{tag}}} - \boldsymbol{p}^*_X - \boldsymbol{p}^*_\ell|$, the absolute value of the missing momentum.
- $\cos\theta^*_{\text{miss}}$, the cosine of the angle between the momentum $\boldsymbol{p}^*_\ell$ of the lepton candidate and the missing momentum $\boldsymbol{p}^*_{\text{miss}}$.
- E^W_{90}, the energy in the hemisphere defined by the direction of the virtual $W^\pm$ in the B meson decay,

$$E^W_{90} = \sum_{i \in X,\ \boldsymbol{p}^*_i \cdot \boldsymbol{p}^*_W > 0} E_i^{\text{ECL}},$$

  where the sum extends over all charged and neutral candidates in the recoiling system X that are in the hemi-

sphere of the $W^{\pm}$, and E^{ECL} corresponds to the energy deposited in the ECL. The momentum of the virtual $W^{\pm}$ is calculated as

$$\boldsymbol{p}^*_W \;=\; \boldsymbol{p}^*_\ell + \boldsymbol{p}^*_\nu \approx\; \boldsymbol{p}^*_\ell + \boldsymbol{p}^*_{\mathrm{miss}} = -\boldsymbol{p}^*_X,$$

where the momentum $\boldsymbol{p}_\nu$ of the neutrino is estimated using the missing momentum $\boldsymbol{p}^*_{\mathrm{miss}}$. In the equation above we assume the B meson to be almost at rest in the $\Upsilon(4S)$ frame, that is $\boldsymbol{p}^*_{B^0_{\mathrm{tag}}} \approx \mathbf{0}$ and thus $\boldsymbol{p}^*_{\mathrm{miss}} \approx -\left(\boldsymbol{p}^*_X + \boldsymbol{p}^*_\ell\right)$.

- $|\cos\theta^*_{\mathrm{T}}|$, the absolute value of the cosine of the angle between the momentum $\boldsymbol{p}^*_\ell$ of the lepton and the thrust axis of the tag-side B meson in the $\Upsilon(4S)$ frame. In general, a thrust axis $\boldsymbol{T}$ can be defined as the unit vector that maximizes the thrust

$$T = \frac{\sum_i |\boldsymbol{T}\cdot\boldsymbol{p}_i|}{\sum_i |\boldsymbol{p}_i|},$$

where the sum extends over a group of particles. For the thrust axis of the tag-side B meson, the sum extends over all charged and neutral candidates on the tag side.

Intermediate electron, intermediate muon, and intermediate kinetic lepton These categories exploit flavor signatures from secondary leptons produced through the decay of charmed mesons and baryons occurring via transitions $b\to\ c\to\ s\ (d)\ \ell^+\ \nu_\ell$. In this case the charge-flavor correspondence is reversed with respect to primary leptons: a positively charged secondary lepton tags a $\bar{B}^0$ meson, and a negatively charged one a B^0 meson. Since their momentum spectrum is much softer in comparison with primary leptons, we refer to secondary leptons as intermediate leptons.

Kaon This category exploits the signature from kaons originating from decays of charmed mesons and baryons produced via $b\to\ c\to\ s$ transitions. Such kaons are referred to as *right-sign* kaons. They tag a $\bar{B}^0$ if they are negatively charged, and a B^0 if they are positively charged.

The kaon category provides the largest tagging power due to the high abundance of charged kaons (around 80% of the B decays contain one) and because the fraction of right-sign kaons (around 70%) is much larger than the fraction of wrong-sign kaons (around 10%) produced through processes of the kind $\bar{b}\to W^+\left(\to c\bar{s}/c\bar{d}\right)X$, with $c\to s\to K^-$.

To identify target kaons, we include the following input variables:

- $n_{K^0_{\mathrm{S}}}$, the number of reconstructed K^0_{S} candidates on the tag side. Charged kaons originating from $b\to\ c\to\ s$ transitions are usually not accompanied by K^0_{S} candidates, while wrong-sign kaons or charged kaons originating from $s\bar{s}$ pairs out of the vacuum are usually accompanied by one or more K^0_{S} candidates.
- $\sum p_{\mathrm{t}}^2$, the sum of the squared transverse momentum of all tracks on the tag side in the lab frame.
- M^2_{rec}, E^W_{90}, p^*_{miss}, $\cos\theta^*_{\mathrm{miss}}$, and $|\cos\theta^*_{\mathrm{T}}|$, the variables in the $\Upsilon(4S)$ frame that discriminate against the lepton background.

Slow pion The target particles of this category are secondary pions from decays $\bar{B}^0\to D^{*+}(\to D^0\pi^+)X^-$. Due to the small mass difference between D^{*+} and D^0, the secondary pions have a soft momentum spectrum and are therefore called slow pions. To identify slow pions we include some variables of the Kinetic Lepton and the kaon category, which help distinguish the background from slow leptons and kaons.

Kaon-pion This category exploits the flavor signatures of decays containing both a right-sign kaon and a slow pion. We use the following input variables to identify both target particles:

- y_{Kaon}, the probability of being a target kaon obtained from the individual Kaon category (see Sect. 4.2).
- y_{SlowPion}, the probability of being a target slow pion obtained from the individual Slow Pion category (see Sect. 4.2).
- $\cos\theta^*_{K\pi}$, the cosine of the angle between the kaon and the slow-pion momentum in the $\Upsilon(4S)$ frame.
- $q_K\cdot q_\pi$, the charge product of the kaon and the slow-pion candidates.

Fast hadron The targets of this category are kaons and pions from the W boson in $b\to\ c\,W^-$ or $b\to\ u\,W^-$ decays, and from *one-prong* decays of primary tau leptons from $b\to\ \tau^-(\to\ h^-\ \nu_\tau)\ \bar{\nu}_\tau\ X$ transitions, where h^- stands for a π^- or a K^-. This category considers as targets also those kaons and pions produced through intermediate resonances that decay via strong processes conserving the flavor information, for example $\bar{B}^0\to K^{*-}(\to K^-\pi^0)\,X^+$. The target kaons and pions are referred to as fast hadrons because of their hard momentum spectrum. To identify them we use the same set of variables as the Slow Pion category, which also distinguish fast kaons and pions among the background of slow particles.

Maximum p* This category is a very inclusive tag based on selecting the charged particle with the highest momentum in the $\Upsilon(4S)$ frame and using its charge as a flavor tag. In this way we give a higher weight to primary particles that may have not been selected either as a primary lepton or as a fast hadron. Primary hadrons and leptons from the $W^{\pm}$ boson in $b\to\ c\ W^-$ or in $b\to\ u\ W^-$ transitions have a very hard

momentum spectrum and are most likely to be the tag-side particles with the largest momenta in a given event.

Fast-slow-correlated (FSC) The targets of this category are both slow pions and high-momentum primary particles. To identify them, we use the following input variables:

- p^*_{Slow}, the momentum of the slow pion candidate in the $\Upsilon(4S)$ frame.
- p^*_{Fast}, the momentum of the fast candidate in the $\Upsilon(4S)$ frame.
- $|\cos\theta^*_{\text{T, Slow}}|$, the absolute value of the cosine of the angle between the thrust axis and the momentum of the slow pion candidate.
- $|\cos\theta^*_{\text{T, Fast}}|$, the absolute value of the cosine of the angle between the thrust axis and the momentum of the fast candidate.
- $\cos\theta^*_{\text{SlowFast}}$, the cosine of the angle between the momenta of the slow and the fast candidate.
- $q_{\text{Slow}} \cdot q_{\text{Fast}}$, the charge product of the slow pion and the fast candidate.

Lambda This category exploits the additional flavor signatures provided by Λ baryons from $b \to c \to s$ transitions. A Λ baryon indicates a $\bar{B}^0$, and a $\bar{\Lambda}$ a B^0. Here, Λ candidates are reconstructed from pairs of proton and pion candidates. To identify target Λ particles, we use the momentum of the reconstructed Λ, the momenta of the proton and the pion, and also the following input variables:

- q_Λ, the flavor of the Λ baryon.
- M_Λ, the reconstructed mass of the Λ.
- $n_{K^0_S}$, the number of reconstructed K^0_S candidates on the tag side.
- $\cos\theta_{\boldsymbol{x}_\Lambda \boldsymbol{p}_\Lambda}$, the cosine of the angle between the Λ momentum $\boldsymbol{p}_\Lambda$ and the direction from the interaction point to the reconstructed Λ vertex $\boldsymbol{x}_\Lambda$ in the lab frame.
- $|\boldsymbol{x}_\Lambda|$, the absolute distance between the Λ vertex and the interaction point.
- σ^{zz}_Λ, the uncertainty on the Λ vertex fit in the direction along the beam (z direction).
- χ^2_Λ, the χ^2 probability of the vertex fit of the reconstructed Λ decay vertex.

In comparison with previous versions of the Belle II flavor taggers [1,31,34], for the current version of the algorithms we exclude track impact parameters (displacement from nominal interaction point), because they are not yet well simulated for small displacements below 0.1 cm. Track impact parameters provide additional separation power between primary particles produced at the B-decay vertex (and thus with small track impact parameters) and secondary particles with decay vertices displaced from the interaction point. Thus, we will consider to use them again in the future.

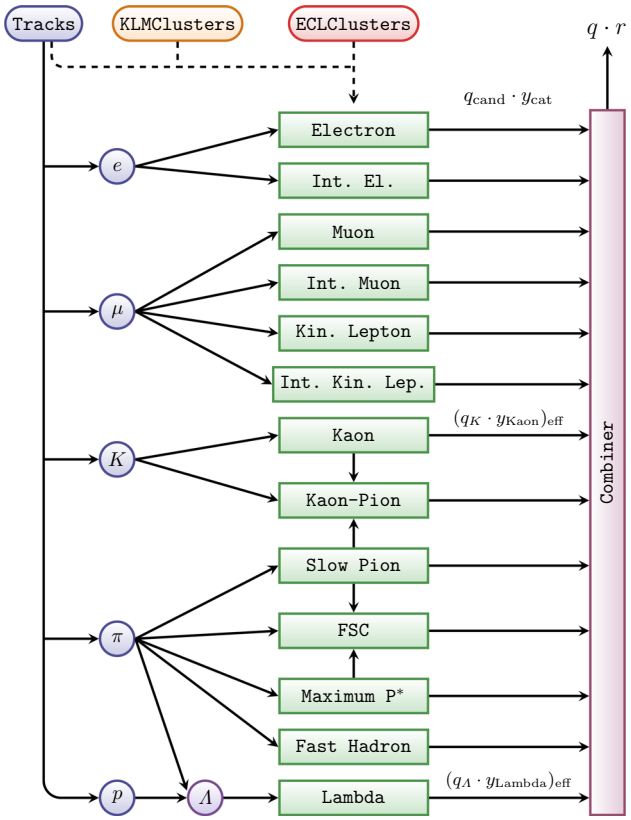


Fig. 1 Schematic overview of the category-based flavor tagger. The tracks on the tag side are used to build five different lists of candidates: e, μ, K, π, and p. Each category considers the list of candidates belonging to its own targets. The different categories are represented by green boxes, and the combiner by a magenta box

For the current version of the algorithms, we also exclude the p-value of the track fit for the Muon and the Kinetic Lepton categories since we observe discrepancies between data and simulation in the p-value distribution of particles identified as primary muons.

4.2 Algorithm

The category-based flavor tagger performs a two-level procedure with an event level for each category followed by a combiner level. Figure 1 shows a schematic overview. The algorithm is based on Fast Boosted Decision Tree (FBDT) [48] classifiers, which are stochastic gradient-boosted decision trees that incorporate several mechanisms for regularization and are optimized to save computing resources during training and application.

At the event level, the flavor tagger identifies decay products providing flavor signatures among the $e^\pm$, $\mu^\pm$, $K^\pm$, $\pi^\pm$, and Λ candidates. Each category considers the list of particle candidates corresponding to its target particles. The event-

level process is performed for each category, which corresponds to an FBDT classifier that receives the input variables associated with the category.

The event-level multivariate method assigns to each particle candidate a real-valued output $y_{\text{cat}} \in [0, 1]$ corresponding to the probability of being the target of the corresponding category providing the right flavor tag. Within each category, the particle candidates are ranked according to the values of y_{cat}. The candidate with the highest y_{cat} is identified as flavor-specific decay product. Figure 2 illustrates the procedure. Only for the Maximum p^* category, the candidates are ranked according to their momenta in the $\Upsilon(4S)$ frame. Two special categories get information from other categories: the Kaon-Pion category and the Fast-Slow-Correlated (FSC) category.

At the combiner level, the algorithm combines the information provided by all categories into the final product $q \cdot r$ using a combiner-level FBDT. Combining the information provided by all categories improves the performance of the flavor tagger as the B decays possibly offer more than one flavor-specific signature. The combiner receives an input from each category corresponding to the product $q_{\text{cand}} \cdot y_{\text{cat}}$, where q_{cand} is the charge (or q_Λ flavor) of the candidate identified as flavor-specific decay product, and y_{cat} is the probability provided by the event-level FBDT. Only for the Kaon and the Lambda categories the input is the effective product

$$(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}} = \frac{\prod_i \left(1 + \left(q_{\text{cand}} \cdot y_{\text{cat}}\right)_i\right) - \prod_i \left(1 - \left(q_{\text{cand}} \cdot y_{\text{cat}}\right)_i\right)}{\prod_i \left(1 + \left(q_{\text{cand}} \cdot y_{\text{cat}}\right)_i\right) + \prod_i \left(1 - \left(q_{\text{cand}} \cdot y_{\text{cat}}\right)_i\right)},$$

where the products extend over the three particles with the highest y_{cat} value. The use of $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ for the Kaon and the Lambda categories slightly improves the tagging performance. We find no significant improvement when we use it for the other categories.

The structure of the FBDT classifiers, the learning procedure, and the preprocessing of the input data is controlled with different so-called *hyper-parameters*. We use the default hyper-parameter values optimized for the Full-Event-Interpretation algorithm [49,50], which performs similarly complex classifications to identify B mesons and other intermediate particles. The *number of levels* in each tree is three; the *number of cuts* for the cumulative probability histograms of each input variable is eight; the fraction of the sample to train each tree (*sampling rate*) is 0.5; the *learning rate* to regulate the training is 0.1. For the flavor tagger, only the *number of trees* was optimized to 500. For the training procedure, the FBDT algorithm transforms the distribution of the input variables to a uniform distribution and uses a negative binomial log-likelihood loss function.

The FBDT algorithm provides an internal ranking of input variables by counting how often the variables are used to split decision tree nodes and by weighting each split according to the separation gain and the number of events in the node [48,49]. Based on this ranking [31], we generally observe that the input variables with largest separation power at the event level are the PID variables followed by the particle momenta. Variables requiring information from all tracks and neutral clusters, for example M^2_{rec}, E^W_{90}, and $\cos\theta^*_{\text{miss}}$, provide marginal additional separation power. At the combiner level, the categories with largest separation power are the Kaon and the Kinetic Lepton categories followed by the Maximum p^*, Slow Pion, FSC, and Fast Hadron categories. The other categories provide marginal additional separation power.

5 The deep-learning flavor tagger

To explore the advantages of deep-learning multivariate methods, we developed a DNN flavor tagger based on a deep-learning multi-layer perceptron (MLP). The algorithm is designed to learn the correlations between the characteristics of the tag-side tracks and the flavor of the tag-side B meson avoiding any pre-selection of decay products. The algorithm provides as output the product $q \cdot r$. The implementation of the algorithm is based on the machine-learning library Tensor-flow [51].

The DNN flavor tagger sorts the tracks on the tag side into two groups, a positive and a negative one, depending on the electric charge of the particle (see Fig. 3). The algorithm ranks the tracks in each group according to their momenta in the $\Upsilon(4S)$ frame, and selects the top five tracks in each group. We find on average around six tag-side tracks per event in simulation with about equal fraction of positive and negative tracks. About 96% of the events have fewer than ten tag-side tracks. If an event contains fewer than five positive or fewer than five negative tracks, the algorithm sets the input variables for the missing candidates to zero.

For each charged-particle candidate, the deep-learning MLP receives ten input variables in the current configuration: five PID likelihoods $\mathcal{L}_e$, $\mathcal{L}_\mu$, $\mathcal{L}_K$, $\mathcal{L}_\pi$, and $\mathcal{L}_p$, the magnitude of the momentum p^*, the cosine of the polar angle $\cos\theta^*$, the azimuth angle ϕ^*, and the number of hits in the vertex detectors N_{PXD} and N_{SVD}. Multiplying the number of input variables by the number of candidates yields 100, corresponding to the number of input nodes.

We optimize the hyper-parameters of the MLP by performing various scans. The optimized MLP contains eight *hidden layers* with 300 nodes each. Based on previous studies on similarly complex classification tasks [52], we employ the tanh function as *activation function* to describe possible

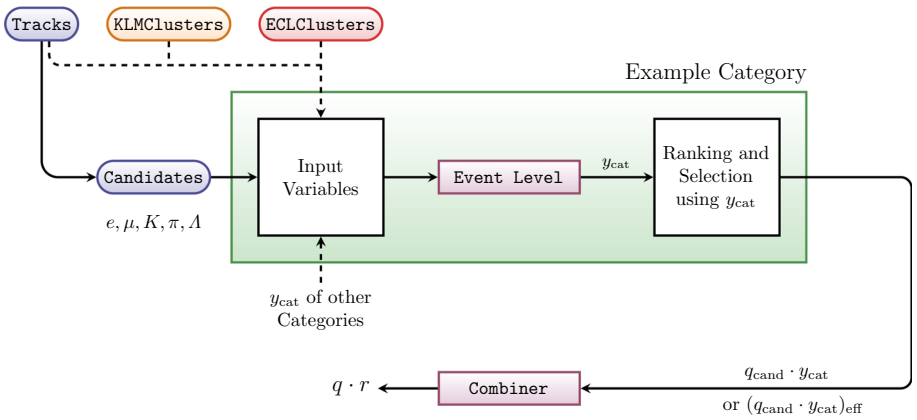


Fig. 2 Procedure for each single category (green box): the candidates correspond to the reconstructed tracks for a specific mass hypothesis. Some of the input variables consider all reconstructed tracks and all neutral ECL and KLM clusters on the tag side. The magenta boxes represent multivariate methods: y_{cat} is the output of the event level. The output of the combiner is equivalent to the product $q \cdot r$. Each box corresponds to an FBDT classifier

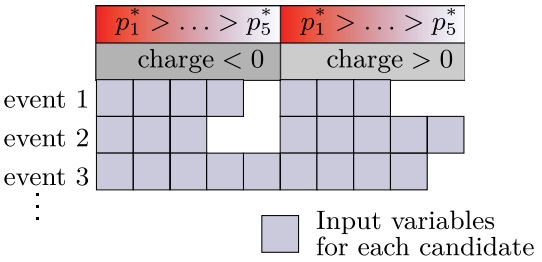


Fig. 3 Scheme of the input information for the MLP in the deep-learning flavor tagger. The tracks on the tag side are sorted into two groups according to their electric charge and ranked according to their momenta in the $\Upsilon(4S)$ frame. The top 5 tracks in each group are selected as candidates. The MLP receives input variables for each candidate

non-linear dependences between the inputs and the B-meson flavor. The inputs are transformed to be uniformly distributed in the range $[-1, 1]$ corresponding to the range of the activation function. Zero, the default value for missing tracks, corresponds to the mean of the transformed distribution.

For the training procedure, we use a binary cross-entropy loss function with regularization terms [34]. The loss function is minimized using a mini-batch stochastic gradient-descent algorithm based on backpropagation [53].

6 Training procedure

Both flavor taggers are trained using $B^0 \to \nu_\tau \bar{\nu}_\tau$ samples (see Sect. 3). In this way, we avoid possible bias due to CP asymmetries or reconstruction performance since these samples are generated without built-in CP violation, and all reconstructed objects (tracks, photons, and KLM clusters) can be used to form the tag side without passing through reconstruction of the signal side.

The algorithms are trained with a sample of about ten million MC events and tested afterwards with an independent sample of the same size to exclude overtraining. We find no significant improvement in tagging performance using two to five times larger training samples. We train the algorithms for each MC campaign to optimize them for the most up-to-date data processing and background expectation.

For the category-based algorithm, the training sample is divided into two statistically independent MC samples of the same size: one sample for the event level, and one sample for the combiner level. The event level is trained first and each category is trained independently. The combiner is trained afterwards.

For the DNN algorithm, we take about 10% of the training sample as an independent validation sample. We monitor the training procedure by calculating the value of the loss function on the validation sample at each training epoch and stop the training procedure if the value starts increasing for a fixed number of 100 epochs. We then save the MLP configuration at the epoch leading to the best performance on the validation sample. Typically, the training is stopped after about 500 epochs. We train 10 different MLPs with different initial random weights and keep only the one leading to the best performance.

Over-fitting is checked for each of the multivariate methods in both flavor taggers by comparing the distribution of the output on the training sample with the output on the testing sample. The output on the training and on the testing sample have to be statistically compatible.

For the DNN tagger, the MLP complexity calls for significant computing resources to train the algorithm. We use GPUs to train the deep-learning MLP to exploit their parallel computation capabilities. On a GTX 970 GPU [54], the training procedure for the eight-layer MLP takes about 48 hours. In comparison, the training procedure for the category-based flavor tagger takes about five hours running on a single CPU core.

We compare the performance of both flavor taggers using the testing $B^0 \to \nu_\tau \bar{\nu}_\tau$ sample. Figure 4 shows the 2D distribution of the DNN output vs. the combiner FBDT output. From the sample we estimate a Pearson correlation coefficient around 90%. Figure 5 shows the receiver operating characteristics (ROC) and the area under the ROC curve (AUC) for all events, for events containing a target particle of the

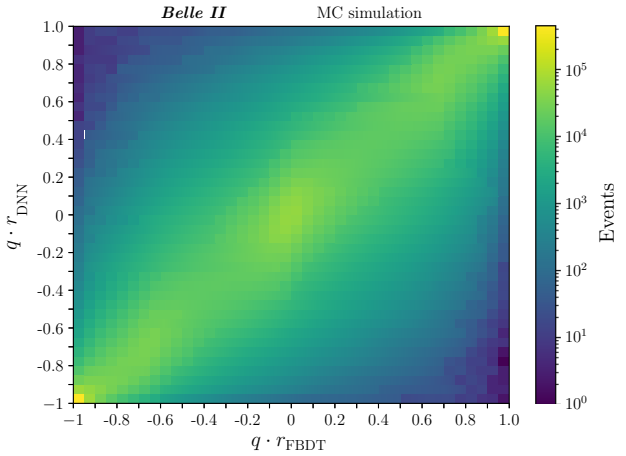


Fig. 4 Distribution of the DNN tagger output vs. the category-based (FBDT) tagger output in the testing $B^0 \to \nu_\tau \bar{\nu}_\tau$ simulation sample

Kinetic Lepton or Kaon categories, and for events containing less than 5 tracks, 5 to 10 tracks, and more than 10 tracks. The category-based tagger reaches a slightly better performance for events with a target of the Kinetic Lepton category and for events with more than ten tracks. On the other hand, the DNN tagger reaches a slightly better performance for events with one target of the Kaon category and for events with less than 10 tracks. However, in general, both algorithms reach about the same performance for all events and for the various sub-samples.

After the training, we perform checks using signal-only MC samples, where the signal B meson decays to one benchmark mode such as $B^0 \to K^+\pi^-$, $B^0 \to J/\psi(\to \mu^+\mu^-)\, K^0_S(\to \pi^+\pi^-)$, $B^0 \to \eta'(\to \pi^+\pi^-\, \eta(\to \pi^+\pi^-\pi^0))\, K^0_S(\to \pi^+\pi^-)$, or one of the neutral B decays listed in the following section. We reconstruct the signal B decay in each event and use the tag-side objects as input for the flavor taggers. For correctly associated MC events, we verify that the tagging performance is consistent with the one obtained using the $B^0 \to \nu_\tau \bar{\nu}_\tau$ sample.

7 Reconstruction of calibration samples

To evaluate the performance of the flavor taggers, we reconstruct the following signal B decays,

- $B^+ \to \bar{D}^0 \pi^+$,
- $B^+ \to \bar{D}^0 \rho^+$,
- $B^+ \to \bar{D}^{*0}(\to \bar{D}^0 \pi^0)\, \pi^+$,
- $B^+ \to \bar{D}^{*0}(\to \bar{D}^0 \pi^0)\, \rho^+$,
- $B^+ \to \bar{D}^{*0}(\to \bar{D}^0 \pi^0)\, a_1^+$,
- $B^0 \to D^- \pi^+$,
- $B^0 \to D^- \rho^+$,
- $B^0 \to D^{*-}(\to \bar{D}^0 \pi^-)\, \pi^+$,
- $B^0 \to D^{*-}(\to \bar{D}^0 \pi^-)\, \rho^+$,
- $B^0 \to D^{*-}(\to \bar{D}^0 \pi^-)\, a_1^+$,

for which we reconstruct the following D decays,

- $\bar{D}^0 \to K^+\pi^-$,
- $\bar{D}^0 \to K^+\pi^-\pi^+\pi^-$,
- $\bar{D}^0 \to K^+\pi^-\pi^0$,
- $\bar{D}^0 \to K^0_S\pi^+\pi^-$,
- $D^- \to K^+\pi^-\pi^-$,
- $D^- \to K^0_S\pi^-$,
- $D^- \to K^0_S\pi^-\pi^0$,
- $D^- \to K^+\pi^-\pi^-\pi^0$.

7.1 Reconstruction and baseline selection

We reconstruct charged-pion and charged-kaon candidates by starting from the most inclusive charged-particle selections. To reduce the background from tracks that do not originate from the interaction region, we require fiducial criteria that restrict the candidates to loose ranges of displacement from the nominal interaction point ($|dr| < 0.5$ cm radial and $|dz| < 3$ cm longitudinal) and to the full polar-acceptance in the central drift chamber ($17^\circ < \theta < 150^\circ$). Additionally, we use PID information to identify kaon candidates by requiring the likelihood $\mathcal{L}_K$ to be larger than 0.4.

We reconstruct neutral pion candidates by requiring photons to exceed energies of 80 MeV in the forward region, 30 MeV in the central volume, and 60 MeV in the backward region. We restrict the diphoton mass to be in the range $120 < M(\gamma\gamma) < 145\,\mathrm{MeV}/c^2$. The mass of the π^0 candidates is constrained to its known value [40] in subsequent kinematic fits.

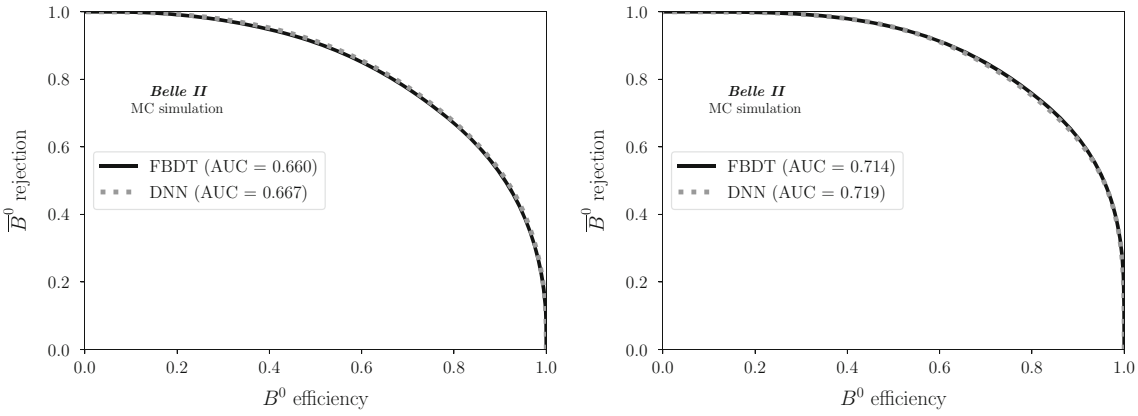
For K^0_S reconstruction, we use pairs of oppositely charged particles that originate from a common decay vertex and have a dipion mass in the range $450 < M(\pi^+\pi^-) < 550\,\mathrm{MeV}/c^2$. To reduce combinatorial background, we apply additional requirements, dependent on K^0_S momentum, on the distance between trajectories of the two charged-pion candidates, the K^0_S flight distance, and the angle between the pion-pair momentum and the K^0_S flight direction.

The resulting $K^\pm$, $\pi^\pm$, π^0, and K^0_S candidates are combined to form $D^{(*)}$ candidates in the various final states, by requiring their invariant masses to satisfy

- $1.84 < M\left(K^+\pi^-,\, K^+\pi^-\pi^+\pi^-,\, K^+\pi^-\pi^0,\, K^0_S\pi^+\pi^-\right) < 1.89\,\mathrm{GeV}/c^2$,
- $1.844 < M\left(K^+\pi^-\pi^-,\, K^0_S\pi^-, K^0_S\pi^-\pi^0,\, K^+\pi^-\pi^-\pi^0\right) < 1.894\,\mathrm{GeV}/c^2$,
- $0.14 < M(D^0\pi^0) - M(D^0) < 0.144\,\mathrm{GeV}/c^2$,
- $0.143 < M(D^0\pi^+) - M(D^0) < 0.147\,\mathrm{GeV}/c^2$,

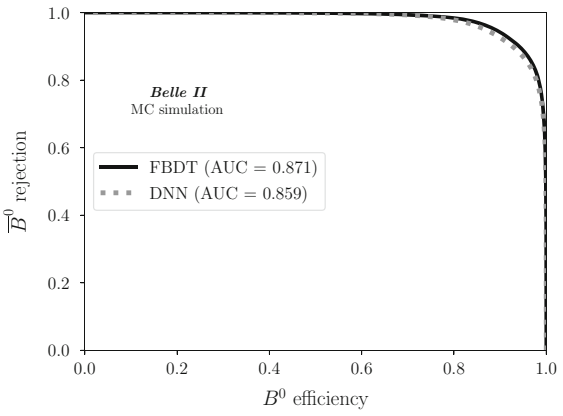
where $M(D^0)$ and $M(D^0\pi^0,\, D^0\pi^+)$ are the invariant masses of the reconstructed D^0 and D^* candidates. We reconstruct $\rho^\pm$ candidates from pairs of charged and neutral pions, and $a_1^\pm$ candidates from three charged pions with the following requirements:

- $|M(\pi^+\pi^0) - M_\rho| < 0.15\,\mathrm{GeV}/c^2$,
- $|M(\pi^+\pi^-\pi^+) - M_{a_1}| < 0.4\,\mathrm{GeV}/c^2$,

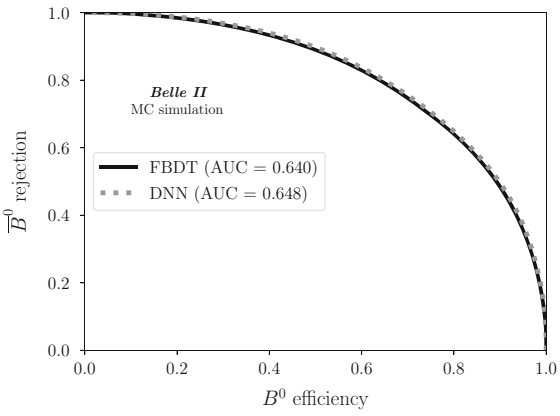


(a) All events

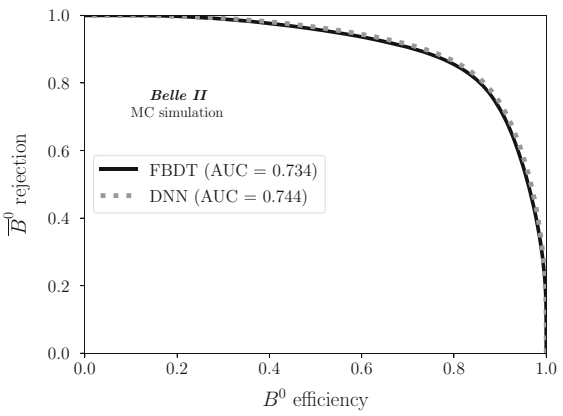
(d) Events with less than 5 tracks



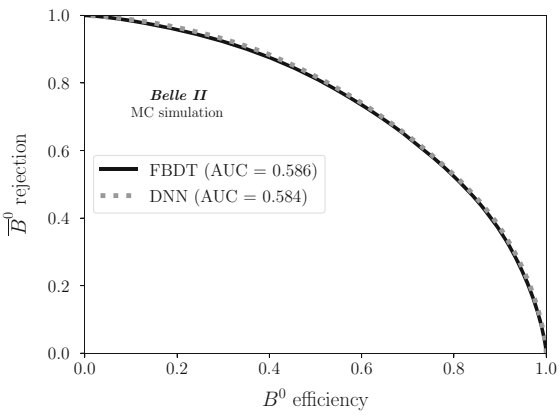
(b) Events with one Kinetic Lepton target



(e) Events with 5 to 10 tracks



(c) Events with one Kaon target



(f) Events with more than 10 tracks

Fig. 5 Receiver operating characteristic curves comparing the performance of the DNN tagger and the category-based (FBDT) tagger in the testing $B^0 \to \nu_\tau \bar{\nu}_\tau$ simulation sample. Curves and values of area under the curve (AUC) are shown for (left) all events, events with one target particle of the Kinetic Lepton category and events with one target particle of the Kaon category, and (right) events with less than 5 tracks, with 5 to 10 tracks and with more than 10 tracks

where M_ρ and M_{a_1} are the known masses [40] of the ρ and a_1 mesons. To identify primary $\pi^\pm$ and $\pi^\pm$ candidates used to reconstruct $\rho^\pm$ and $a_1^\pm$ candidates, we also require the likelihood $\mathcal{L}_\pi$ to be larger than 0.1 and the $\pi^\pm$ momentum in the $\Upsilon(4S)$ frame to be larger than $0.2\,\mathrm{GeV}/c$.

To reconstruct the signal B candidates, we combine the $D^{(*)}$ candidates with appropriate additional candidate particles, $\pi^\pm$, $\rho^\pm$ or $a_1^\pm$, by performing simultaneous kinematic-vertex fits of the entire decay chain [55] into each of our signal channels. We perform the kinematic-vertex fits without constraining the decay-vertex position or the invariant mass of the decaying particles and require the fit to converge. Requiring the kinematic-vertex fit to converge keeps about 96% of the correctly associated MC events.

We use the following kinematic variables to distinguish B signals from the dominant continuum background from $e^+e^- \to u\bar{u}$, $d\bar{d}$, $s\bar{s}$, and $c\bar{c}$ processes:

- $M_{\mathrm{bc}} \equiv \sqrt{s/(4c^4) - (p_B^*/c)^2}$, the beam-energy constrained mass, which is the invariant mass of the B candidate calculated with the B energy replaced by half the collision energy $\sqrt{s}$, which is more precisely known;
- $\Delta E \equiv E_B^* - \sqrt{s}/2$, the energy difference between the energy E_B^* of the reconstructed B candidate and half of the collision energy, both measured in the $\Upsilon(4S)$ frame.

We retain B candidates that have $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ and $|\Delta E| < 0.12\,\mathrm{GeV}$. Additionally, for channels with $\rho^\pm$ candidates, we remove combinatorial background from soft π^0 mesons collinear with the $\rho^\pm$, by requiring the cosine of the helicity angle θ_{H} between the B and the π^+ momenta in the ρ frame to satisfy $\cos\theta_{\mathrm{H}} < 0.8$.

We form the tag side of the signal B candidates using all remaining tracks and photons that fulfill the loose fiducial criteria, and KLM clusters. The category-based and the DNN taggers receive the tag-side objects and run independently of each other.

7.2 Continuum suppression and final selection

To suppress continuum background, we apply requirements on the two topological variables with the highest discrimination power between signal from hadronic B decays and continuum background: $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$, the cosine of the angle between the thrust axis of the signal B (reconstructed) and the thrust axis of the tag-side B (remaining tracks and clusters), and R_2, the ratio between the second and zeroth Fox-Wolfram moments [56] calculated using the full event information.

We vary the selections on $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$ and R_2 to maximize the figure of merit $\mathrm{S}/\sqrt{\mathrm{S}+\mathrm{B}}$, where S and B are the number of signal and background B candidates in the signal-enriched range $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ and $-0.12 < \Delta E < 0.09\,\mathrm{GeV}$. Both $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$ and R_2 requirements are optimized simultaneously using simulation. We optimize the requirements for charged and for neutral candidates independently. The optimized requirements are found to be $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}} < 0.87$ and $R_2 < 0.43$ for charged B candidates, and $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}} < 0.95$ and $R_2 < 0.35$ for neutral B candidates.

Applying the optimized R_2 and $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$ requirements keeps about 81% of the charged signal B candidates and about 77% of the neutral ones, and improves the figure of merit by about 12% for charged B candidates and by about 14% for neutral ones. We observe no significant difference in tagging performance before and after the R_2 and $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$ requirements.

After applying the $\cos\theta_{\mathrm{T}}^{\mathrm{sig,tag}}$ and R_2 requirements, more than one candidate per event populate the resulting ΔE distributions, with average multiplicities for the various channels ranging from 1.00 to 3.00 (about 75% of the channels have multiplicities between 1.00 and 2.00). We select a single B candidate per event by selecting the one with the highest p-value of the kinematic-vertex fit. The analyses of charged and neutral B channels are independent: we select one candidate among the charged and one among the neutral channels independently.

Table 3 Tagging efficiencies $\varepsilon \pm \delta\varepsilon$ for charged and neutral $B \to D^{(*)}h^+$ candidates in data and in simulation. All values are given in percent. The uncertainties are only statistical

Channel	MC	Data
$B^0 \to D^{(*)-}h^+$	99.77 ± 0.01	99.75 ± 0.02
$B^+ \to \bar{D}^{(*)0}h^+$	99.80 ± 0.01	99.73 ± 0.02

8 Determination of efficiencies and wrong-tag fractions

The tagging efficiency ε corresponds to the fraction of events to which a flavor tag can be assigned. Since the category-based and the DNN algorithms need only one charged track on the tag side to provide a tag, the tagging efficiency is close to 100% for both, with good consistency between data and simulation as Table 3 shows.

To estimate the fraction of wrongly tagged events w, we fit the time-integrated $B^0 - \bar{B}^0$ mixing probability to the data. We take into account that ε and w can be slightly different for B^0 and $\bar{B}^0$ mesons due to charge-asymmetries in detection and reconstruction. We express ε and w as

$$\varepsilon = \frac{\varepsilon_{B^0} + \varepsilon_{\bar{B}^0}}{2}, \qquad w = \frac{w_{B^0} + w_{\bar{B}^0}}{2},$$

and introduce the differences

$$\Delta\varepsilon = \varepsilon_{B^0} - \varepsilon_{\bar{B}^0}, \quad \Delta w = w_{B^0} - w_{\bar{B}^0},$$

where the subscript corresponds to the true flavor, for example w_{B^0} is the fraction of true B^0 mesons that are wrongly classified as $\bar{B}^0$.

For neutral $B\bar{B}$ pairs produced at the $\Upsilon(4S)$, the time-integrated probability for an event with a signal B flavor $q_{\rm sig} \in \{-1, +1\}$ and tag-side B flavor $q_{\rm tag} \in \{-1, +1\}$ is given by

$$\mathcal{P}(q_{\rm sig}, q_{\rm tag}) = \frac{1}{2}\Big[1 - q_{\rm sig} \cdot q_{\rm tag} \cdot (1 - 2 \cdot \chi_d)\Big],$$

where χ_d is the time-integrated $B^0 - \bar{B}^0$ mixing probability, whose current world average is $\chi_d = 0.1858 \pm 0.0011$ [57]. The equation above assumes that for any event the signal and the tag-side B flavor are correctly identified. To include the effect of the flavor tagging algorithms, one can express the observed probability $\mathcal{P}(q_{\rm sig}, q_{\rm tag})^{\rm obs}$ in terms of the efficiencies ε_{B^0} and $\varepsilon_{\bar{B}^0}$, and the wrong tag fractions w_{B^0} and $w_{\bar{B}^0}$. The probability becomes

$$\begin{aligned}
\mathcal{P}^{\rm obs}(q_{\rm sig}, q_{\rm tag} = +1) &= \varepsilon_{B^0}(1 - w_{B^0}) \cdot \mathcal{P}(q_{\rm sig}, q_{\rm tag} = +1) \\
&\quad + \varepsilon_{\bar{B}^0} w_{\bar{B}^0} \cdot \mathcal{P}(q_{\rm sig}, q_{\rm tag} = -1), \\
\mathcal{P}^{\rm obs}(q_{\rm sig}, q_{\rm tag} = -1) &= \varepsilon_{\bar{B}^0}(1 - w_{\bar{B}^0}) \cdot \mathcal{P}(q_{\rm sig}, q_{\rm tag} = -1) \\
&\quad + \varepsilon_{B^0} w_{B^0} \cdot \mathcal{P}(q_{\rm sig}, q_{\rm tag} = +1),
\end{aligned}$$

which can be written in terms of ε, w, $\mu = \Delta\varepsilon/(2\varepsilon)$ and Δw as

$$\begin{aligned}
\mathcal{P}(q_{\rm sig}, q_{\rm tag})^{\rm obs} = \frac{1}{2}\varepsilon\Big[1 - q_{\rm tag} \cdot \Delta w + q_{\rm tag} \cdot \mu \cdot (1 - 2w) \\
- \big[q_{\rm tag} \cdot (1 - 2w) + \mu \cdot (1 - q_{\rm tag} \cdot \Delta w)\big] \cdot q_{\rm sig} \cdot (1 - 2 \cdot \chi_d)\Big]. \quad (1)
\end{aligned}$$

We sort the events in bins of the dilution factor r provided by the flavor taggers and measure the value of ε, w, μ, and Δw in each r bin (7 bins in total). To compare with our predecessor experiment, we use the binning introduced by Belle [32,33].

Since we need to consider the background, we develop a statistical model with a signal and a background component. We determine the signal yield $N_{\rm sig}$, the background yield $N_{\rm bkg}$, the partial efficiencies ε_i, the wrong-tag fractions w_i, and the asymmetries μ_i and Δw_i in each r-bin i from an extended maximum likelihood fit to the unbinned distributions of ΔE, $q_{\rm sig}$, and $q_{\rm tag}$. We check that the ΔE distribution is statistically independent from those of $q_{\rm sig}$ and $q_{\rm tag}$ with Pearson correlation coefficients below 2%.

In the fit model, the probability density function (PDF) for each component j is given by

$$\mathcal{P}_j(\Delta E, q_{\rm sig}, q_{\rm tag}) \equiv \mathcal{P}_j(\Delta E) \cdot \mathcal{P}_j^{\rm obs}(q_{\rm sig}, q_{\rm tag}).$$

We model the signal ΔE PDF using a Gaussian plus a Crystal Ball function [58] determined empirically using signal MC events obtained from the generic simulation (see Sect. 3), with the additional flexibility of a global shift of peak position and a global scaling factor for the width as suggested by a likelihood-ratio test. The background ΔE PDF is modeled using an exponential function with a floating exponent. Residual peaking backgrounds in generic simulation have expected yields below 0.5% of the signal one and are thus neglected.

The flavor PDF $\mathcal{P}(q_{\rm sig}, q_{\rm tag})^{\rm obs}$ has the same form for signal and background (Eq. 1) with independent ε_i, w_i, Δw_i, μ_i, and χ_d parameters for signal and background. We fix the background $\chi_d^{\rm bkg}$ parameter to zero as we obtain values compatible with zero when we let it float. We find, on the other hand, that the background parameters $\varepsilon_i^{\rm bkg}$, $w_i^{\rm bkg}$, $\Delta w_i^{\rm bkg}$, and $\mu_i^{\rm bkg}$ have to be free to obtain unbiased results for the signal ones.

The total extended likelihood is given by

$$\mathcal{L} \equiv \prod_i \frac{e^{-\sum_j N_j \cdot \varepsilon_i}}{N^i!} \prod_{k=1}^{N^i} \sum_j N_j \cdot \mathcal{P}_j^i(\Delta E^k, q_{\rm sig}^k, q_{\rm tag}^k),$$

where i extends over the r bins, k extends over the events in the r bin i, and j over the two components: signal and background. The PDFs for the different components have no common parameters. Here, N_j denotes the yield for the component j, and N^i denotes the total number of events in the i-th r bin. The partial efficiencies ε_i are included in the flavor part of $\mathcal{P}_j$. Since we can fit only to events with flavor information, the sum of all ε_i must be one. We therefore replace the epsilon for the first bin (with lowest r) with

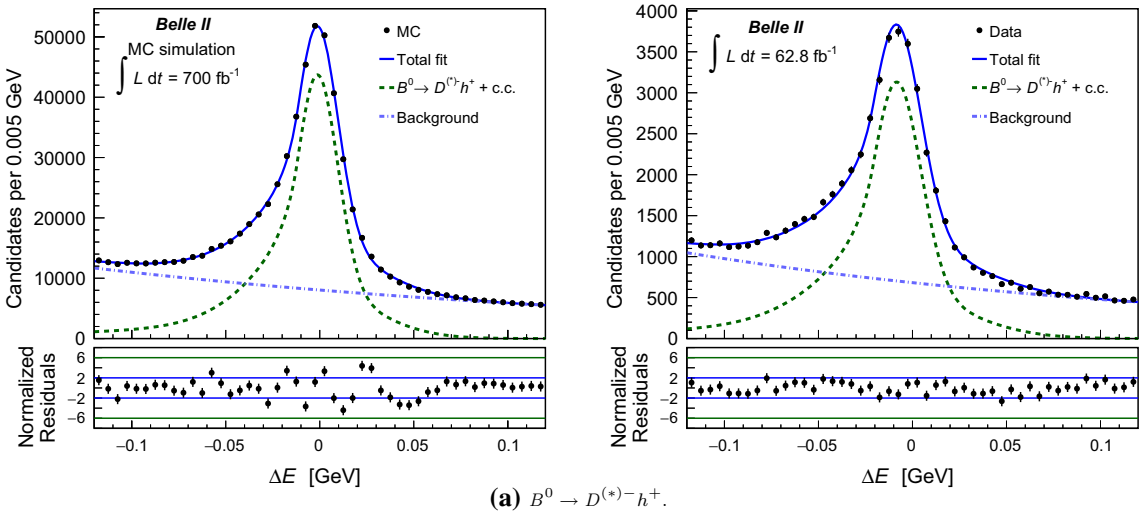
$$\varepsilon_1 = 1 - \sum_{i=2}^{7} \varepsilon_i,$$

and obtain its uncertainty $\delta\varepsilon_1$ from the width of the residuals of simplified simulated experiments.

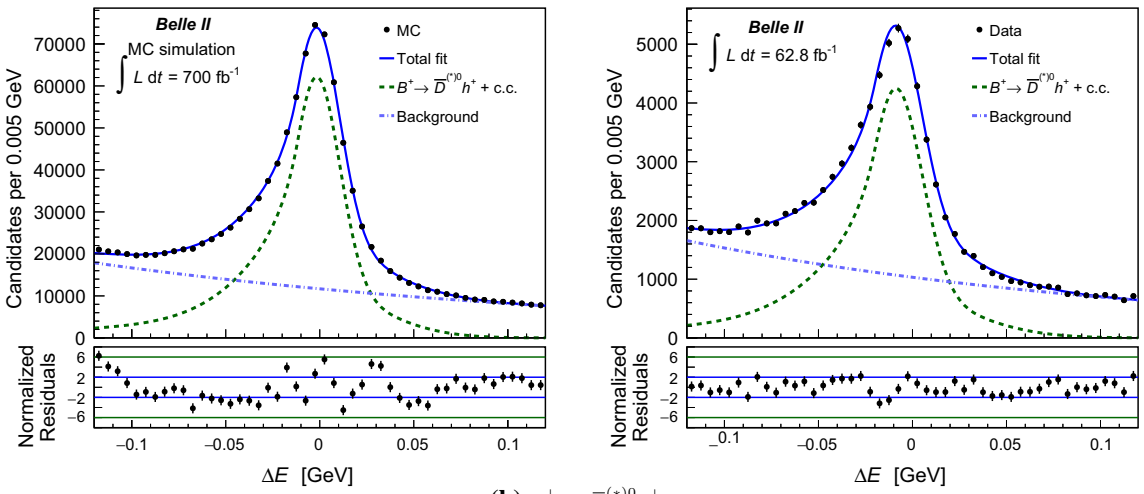
To validate the ΔE model, we first perform an extended maximum likelihood fit to the unbinned distribution of ΔE (without flavor part) in simulation and data. Figure 6 shows the ΔE fit projections in data and simulation for charged and neutral $B \to D^{(*)}h^+$ candidates. Table 4 summarizes the yields obtained from the fits. We observe a relatively good agreement between data and simulation, but a tendency to lower yields with respect to the expectation, especially for charged signal B candidates.

To determine the partial efficiencies ε_i and the wrong-tag fractions w_i, we perform a fit of the full model in a single step. For neutral candidates, we constrain the value of the signal $\chi_d^{\rm sig}$ parameter via a Gaussian constraint,

$$\mathcal{L} \Rightarrow \mathrm{G}(\chi_d^{\rm sig} - \chi_d, \delta\chi_d) \cdot \mathcal{L},$$



(a) $B^0 \to D^{(*)-}h^+$.

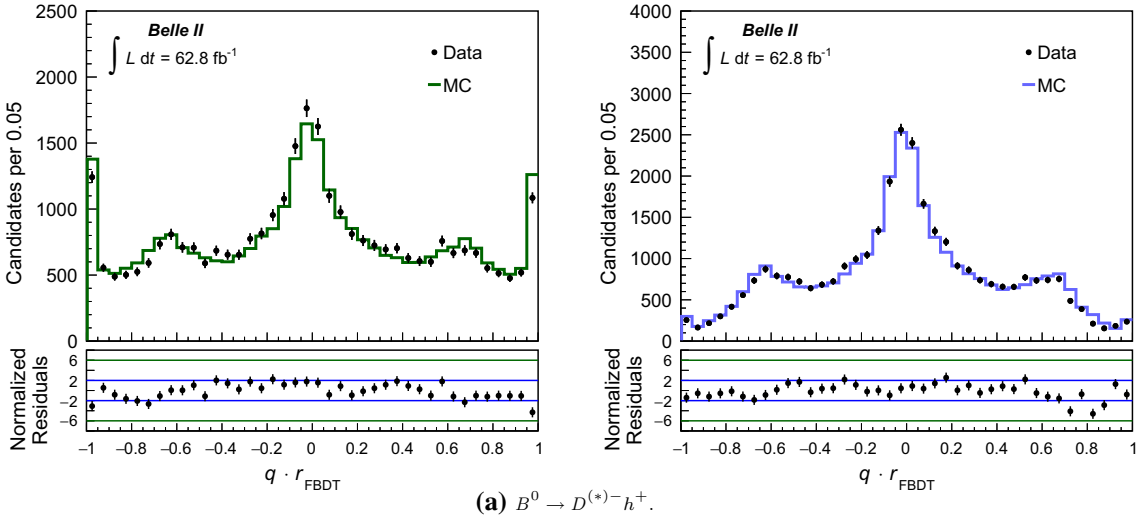


(b) $B^+ \to \bar{D}^{(*)0}h^+$.

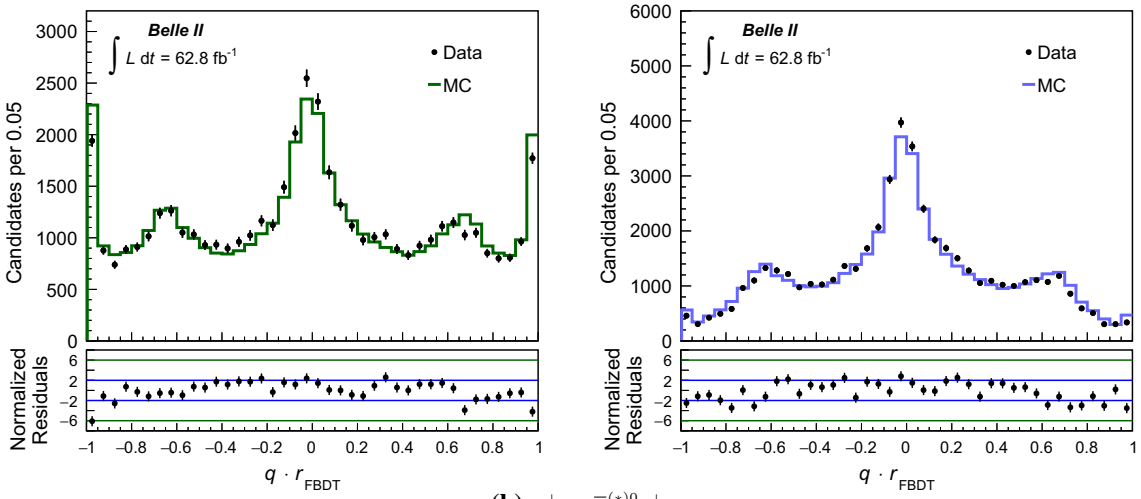
Fig. 6 Distributions of ΔE for (top) neutral and (bottom) charged $B \to D^{(*)}h^+$ candidates reconstructed in (left) simulation and (right) data, restricted to $M_{\rm bc} > 5.27\,\text{GeV}/c^2$. The fit projection of the maximum likelihood fit is overlaid

Table 4 Summary of yields and yields per integrated luminosity obtained from the fit to MC simulation ($700\,\text{fb}^{-1}$) and data ($62.8\,\text{fb}^{-1}$). The uncertainties are only statistical

	Yield		Yield/fb^{-1}	
	MC	Data	MC	Data
$B^0 \to D^{(*)-}h^+$				
Signal	375105 ± 1084	31423 ± 360	536 ± 2	500 ± 6
Background	397275 ± 1094	33764 ± 364	568 ± 2	538 ± 6
$B^+ \to \bar{D}^{(*)0}h^+$				
Signal	610304 ± 1461	46530 ± 482	872 ± 2	741 ± 8
Background	587618 ± 1453	51420 ± 487	840 ± 2	819 ± 8



(a) $B^0 \to D^{(*)-}h^+$.



(b) $B^+ \to \overline{D}^{(*)0}h^+$.

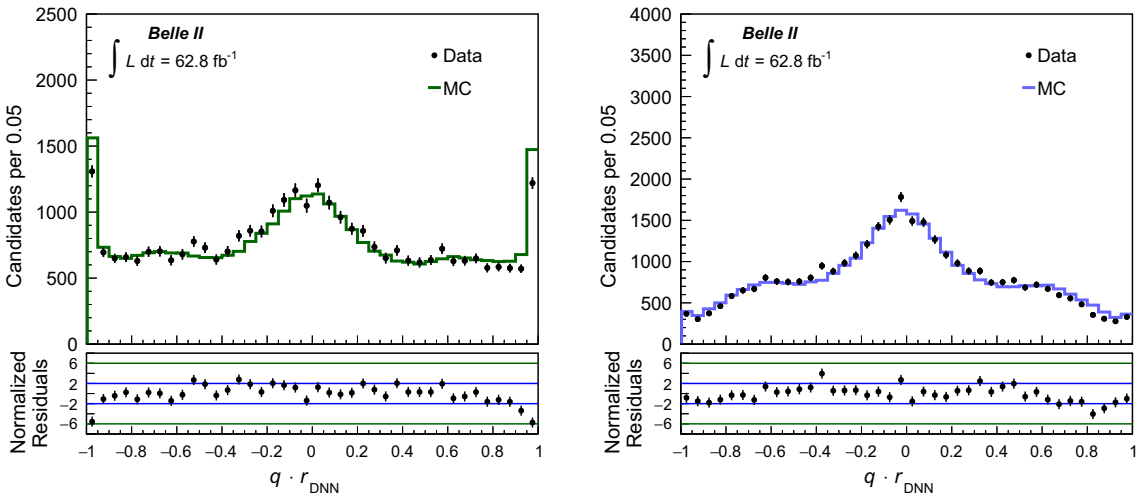
Fig. 7 Normalized $q \cdot r$ distributions obtained with the category-based tagger in data and MC simulation. The contribution (left) from the signal component in data is compared with correctly associated signal MC events and (right) from the background component in data is compared with sideband MC events for (top) neutral and (bottom) charged $B \to D^{(*)}h^+$ candidates

where χ_d and $\delta\chi_d$ are the central value and the uncertainty of the world average. For charged B mesons, χ_d is zero since there is no flavor mixing.

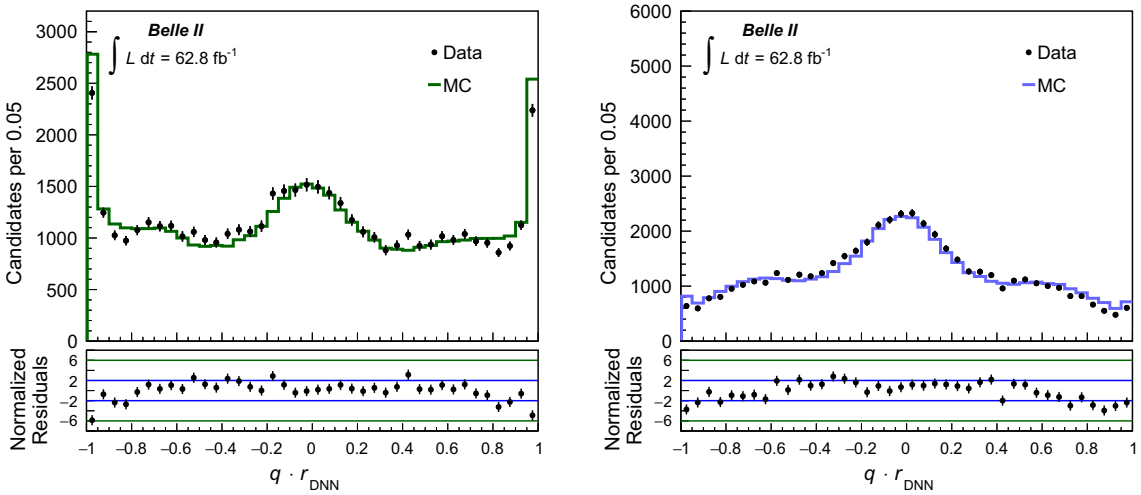
9 Comparison of performance in data and simulation

We check the agreement between data and MC distributions of the flavor-tagger output by performing an $s\mathcal{P}lot$ [59] analysis using ΔE as the control variable. We determine $s\mathcal{P}lot$ weights using the ΔE fit model introduced in the previous section. We weight the data with the $s\mathcal{P}lot$ weights to obtain the individual distributions of the signal and background components in data and compare them with MC simulation. We normalize the simulated samples by scaling the total number of events to those observed in data. The procedure is validated by performing the $s\mathcal{P}lot$ analysis using MC simulation and verifying that the obtained signal and background distributions correspond to the distributions obtained using the MC truth.

Figures 7 and 8 show the $q \cdot r$ distributions provided by the FBDT and by the DNN flavor tagger; the signal and background distributions for neutral and charged $B \to D^{(*)}h^+$ candidates are shown separately. We compare the distribution of the signal component in data with the distribution of correctly associated MC events, and the distribution of the background component in data with the distribution of sideband MC events ($M_{\rm bc} < 5.27\,{\rm GeV}/c^2$ and same fit range



(a) $B^0 \to D^{(*)-}h^+$.



(b) $B^+ \to \overline{D}^{(*)0}h^+$.

Fig. 8 Normalized $q \cdot r$ distributions obtained with the DNN tagger in data and MC simulation. The contribution (left) from the signal component in data is compared with correctly associated signal MC events and (right) from the background component in data is compared with sideband MC events for (top) neutral and (bottom) charged $B \to D^{(*)}h^+$ candidates

$|\Delta E| < 0.12\,\text{GeV}$). We also compare the distributions of the signal component in data with the distribution of correctly associated MC events for the individual tagging categories (Figs. 9, 10, 11). On the MC distributions, the statistical uncertainties are very small and thus not visible.

In general, the results show a good consistency between data and simulation, with a slightly worse performance in data. In the signal $q \cdot r$ distributions, we observe some considerable differences around $|q \cdot r| \approx 1$. We attribute these differences to discrepancies between data and simulation for some of the discriminating input variables, in particular for the electron and muon PID likelihoods. Some differences are observed for categories associated with intermediate slow particles. However, these categories provide only marginal tagging power without degrading the overall tagging performance.

10 Results

We obtain the partial tagging efficiencies ε_i, the wrong-tag fractions w_i, the asymmetries μ_i and Δw_i and the correlation coefficients between them from the maximum-likelihood fit of the full model to data. To evaluate the tagging performance, we calculate the total effective efficiency as

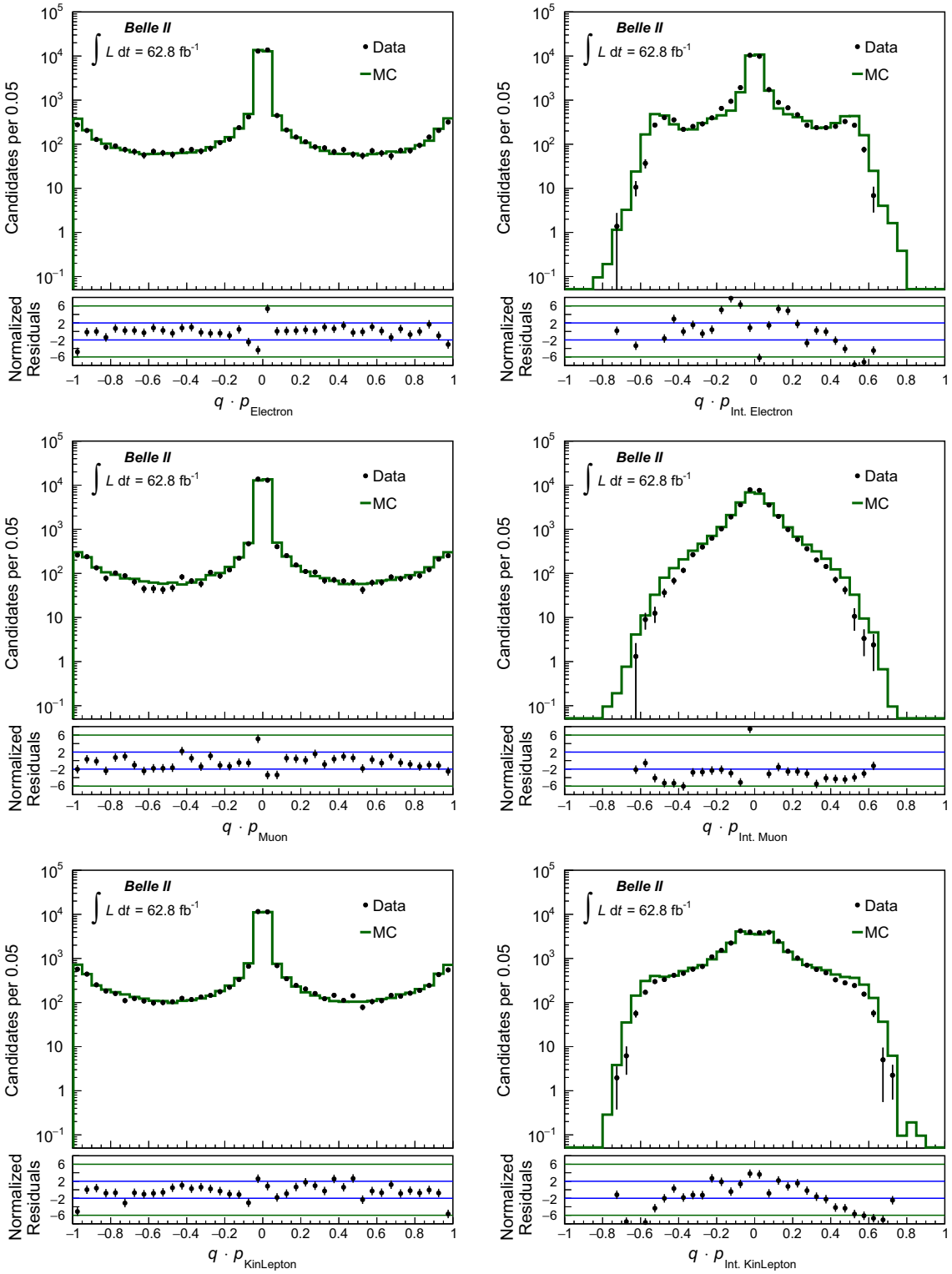


Fig. 9 Normalized output distributions of the Electron, Intermediate Electron, Muon, Intermediate Muon, Kinetic Lepton, and Intermediate Kinetic Lepton categories in data and MC simulation for $B^0 \to D^{(*)-}h^+$ candidates. The contribution from the signal component in data is compared with correctly associated signal MC events

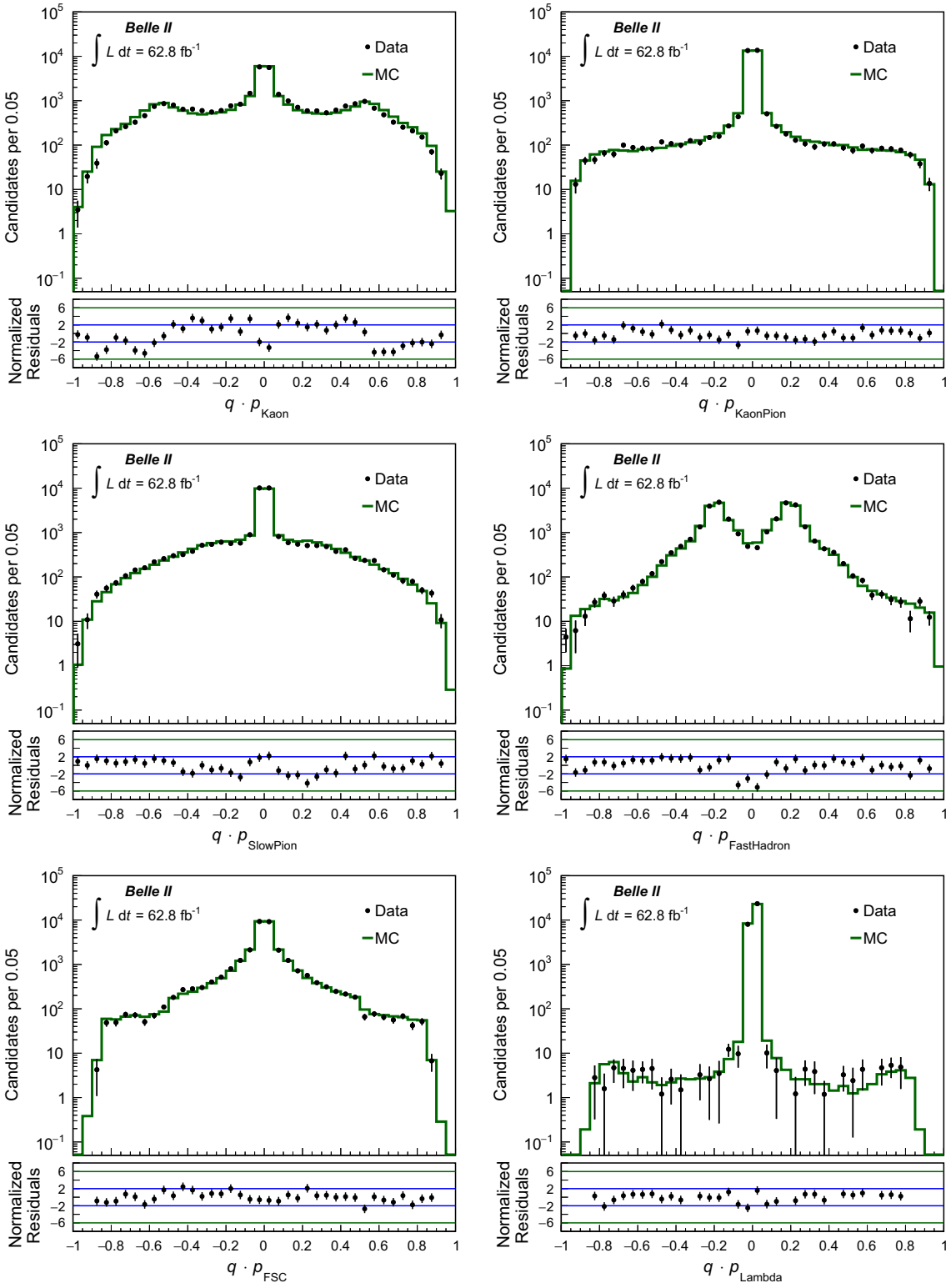


Fig. 10 Normalized output distributions of the Kaon, Kaon-Pion, Slow Pion, Fast Hadron, Fast-Slow-Correlated, and Lambda categories in data and MC simulation for $B^0 \to D^{(*)-}h^+$ candidates. The contribution from the signal component in data is compared with correctly associated signal MC events

$$\varepsilon_{\text{eff}} = \sum_i \varepsilon_{\text{eff},i} = \sum_i \varepsilon_i \cdot (1-2w_i)^2,$$

where $\varepsilon_{\text{eff},i}$ is the partial effective efficiency in the i-th r bin. The effective tagging efficiency is a measure for the effective reduction of events due to the flavor dilution r. In CP-violation analyses, the statistical uncertainty of measured CP asymmetries is approximately proportional to $1/\sqrt{N_{\text{eff}}} = 1/\sqrt{N \cdot \varepsilon_{\text{eff}}}$, where N_{eff} is the number of effectively tagged events. Thus, one would obtain the same statistical precision for N_{eff} perfectly tagged events or for N events tagged with an effective efficiency ε_{eff}.

Tables 5 and 6 show the fit results for the category-based and the DNN flavor taggers. The respective effective efficiencies for both flavor taggers are shown in Tables 7 and 8. Figure 12 shows the Pearson correlation coefficients obtained from the Hessian matrix determined by the fit. We observe considerable dependencies among the ε_i efficiencies for both charged and neutral B candidates, and among the asymmetries Δw_i and μ_i for neutral B candidates.

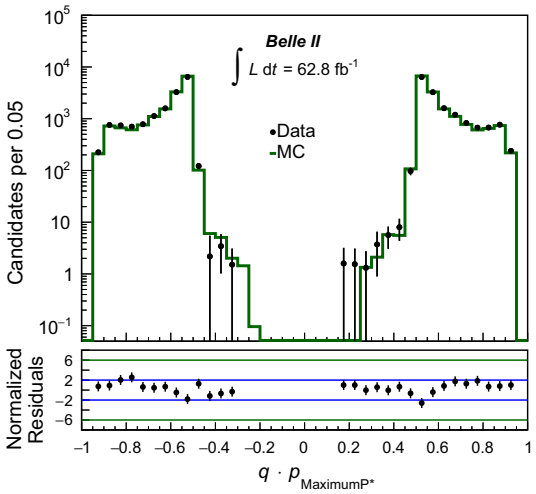


Fig. 11 Normalized output distributions of the Maximum p^* category in data and MC simulation for $B^0 \to D^{(*)-}h^+$ candidates. The contribution from the signal component in data is compared with correctly associated signal MC events

10.1 Systematic uncertainties

We consider the systematic uncertainties associated with the ΔE PDF parametrization, the flavor mixing of the background, the fit bias, and the eventual bias introduced by model assumptions.

ΔE PDF parametrization We perform simplified simulated experiments using an alternative model with a different ΔE parametrization. We perform fits to simulated data samples bootstrapped (sampled with replacement) [60] from the generic MC simulation. We fit using default and alternative models and calculate for each fit parameter x_i the difference δx_i between the results obtained with the alternative model and the results obtained with the default model. We obtain the mean difference $\delta\hat{x}_i$ by fitting a Gaussian function to the distribution of δx_i and take the full mean $\delta\hat{x}_i$ as systematic uncertainty.

For the alternative signal ΔE PDF, we use a triple Gaussian function with the additional flexibility of a global shift of peak position and a global scaling factor for the width. For the alternative background ΔE PDF, we use a second-order Chebyshev polynomial function; we determine the coefficient of the quadratic term by fitting to the distribution of the generic simulation without signal MC events, and then leave the coefficient of the linear term free, in order to have the same degrees of freedom as the default model.

We also check whether the signal ΔE PDF shape changes as a function of r and B flavor and find no significant dependences.

Background mixing Our fit takes into account the uncertainty on the world average for the signal χ_d in the Gaussian constraint. However, we assume that there is no mixing in the background ($\chi_d^{\text{bkg}} = 0$). Since the background includes $B^0\bar{B}^0$ events, we study the effect of flavor mixing in the background by varying the value of the background χ_d^{bkg} by a small amount $\pm\delta\chi_d^{\text{bkg}}$, corresponding to the statistical uncertainty when χ_d^{bkg} is a free parameter in the fit. We find that the difference between the results for $\chi_d^{\text{bkg}} + \delta\chi_d^{\text{bkg}}$ and for $\chi_d^{\text{bkg}} - \delta\chi_d^{\text{bkg}}$ are below 1% of the statistical uncertainty for all fit parameters, which is negligible.

Fit bias For each fit parameter x_i, we determine the fit bias using the residuals from bootstrapped simulated experiments. The residuals are the differences between the fit results for the individual bootstrapped data samples and the fit results for the parent MC sample. We take the full bias as systematic uncertainty.

Fit model Before performing the fit to data, we check that the results of the fit to the full MC sample and the true values determined using MC information agree within about one standard deviation for all fit parameters without tendency of over or underestimation across neighboring r bins. We consider possible bias that cannot be resolved with the current sample sizes and that cannot be attributed to a single effect, for example bias due to fit model or due to MC association, by assigning the uncertainty of the fit to the full MC sample as systematic uncertainty.

We further study possible asymmetries in the reconstruction of signal B candidates that might cause bias in the measurement of the flavor tagging parameters as they are neglected in the fit model. We observe small reconstruction asymmetries between signal $B^0(B^+)$ and $\bar{B}^0(B^-)$ in some individual r-bins below or around 5%. However, we find that

Table 5 Results for ε_i, w_i, μ_i, and Δw_i for the category-based (FBDT) tagger: systematic uncertainties associated with the ΔE parametrization, fit bias, and fit model, and total systematic uncertainty are shown together with the fit results (with stat. and syst. uncertainty). The results are given in percent

Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
$B^0 \to D^{(*)-}h^+$					
ε_1	0.01	0.00	0.09	0.1	$19.0 \pm 0.3 \pm 0.1$
ε_2	0.01	0.00	0.09	0.1	$17.1 \pm 0.3 \pm 0.1$
ε_3	0.01	0.03	0.09	0.1	$21.3 \pm 0.3 \pm 0.1$
ε_4	0.01	0.01	0.07	0.1	$11.3 \pm 0.3 \pm 0.1$
ε_5	0.01	0.01	0.07	0.1	$10.7 \pm 0.3 \pm 0.1$
ε_6	0.00	0.01	0.06	0.1	$8.2 \pm 0.2 \pm 0.1$
ε_7	0.02	0.00	0.07	0.1	$12.4 \pm 0.2 \pm 0.1$
w_1	0.03	0.03	0.45	0.5	$47.1 \pm 1.6 \pm 0.5$
w_2	0.01	0.01	0.46	0.5	$41.3 \pm 1.7 \pm 0.5$
w_3	0.07	0.04	0.40	0.4	$30.3 \pm 1.4 \pm 0.4$
w_4	0.17	0.18	0.53	0.6	$22.9 \pm 1.8 \pm 0.6$
w_5	0.01	0.01	0.49	0.5	$12.4 \pm 1.8 \pm 0.5$
w_6	0.02	0.02	0.50	0.5	$9.4 \pm 1.9 \pm 0.5$
w_7	0.01	0.01	0.37	0.4	$2.3 \pm 1.3 \pm 0.4$
μ_1	0.01	0.03	0.91	0.9	$4.4 \pm 3.2 \pm 0.9$
μ_2	0.15	0.13	0.93	0.9	$3.9 \pm 3.3 \pm 0.9$
μ_3	0.05	0.00	0.82	0.8	$6.8 \pm 2.9 \pm 0.8$
μ_4	0.04	0.01	1.12	1.1	$3.2 \pm 4.0 \pm 1.1$
μ_5	0.16	0.23	1.06	1.1	$-0.5 \pm 4.1 \pm 1.1$
μ_6	0.02	0.04	1.14	1.1	$10.8 \pm 4.3 \pm 1.1$
μ_7	0.37	0.27	0.86	1.0	$-3.7 \pm 3.2 \pm 1.0$
Δw_1	0.16	0.12	0.57	0.6	$8.8 \pm 2.0 \pm 0.6$
Δw_2	0.12	0.15	0.59	0.6	$6.1 \pm 2.1 \pm 0.6$
Δw_3	0.12	0.11	0.54	0.6	$2.7 \pm 1.9 \pm 0.6$
Δw_4	0.05	0.01	0.77	0.8	$5.5 \pm 2.6 \pm 0.8$
Δw_5	0.05	0.03	0.74	0.7	$0.7 \pm 2.9 \pm 0.7$
Δw_6	0.08	0.07	0.84	0.9	$7.7 \pm 3.2 \pm 0.9$
Δw_7	0.19	0.17	0.66	0.7	$0.6 \pm 2.4 \pm 0.7$
$B^+ \to \bar{D}^{(*)0}h^+$					
ε_1	0.04	0.00	0.07	0.1	$18.3 \pm 0.3 \pm 0.1$
ε_2	0.02	0.00	0.07	0.1	$15.4 \pm 0.3 \pm 0.1$
ε_3	0.02	0.01	0.07	0.1	$20.2 \pm 0.3 \pm 0.1$
ε_4	0.00	0.01	0.06	0.1	$11.5 \pm 0.2 \pm 0.1$
ε_5	0.00	0.00	0.06	0.1	$11.9 \pm 0.2 \pm 0.1$
ε_6	0.02	0.00	0.05	0.1	$8.9 \pm 0.2 \pm 0.1$
ε_7	0.07	0.01	0.06	0.1	$13.8 \pm 0.2 \pm 0.1$
w_1	0.01	0.01	0.23	0.2	$48.2 \pm 0.9 \pm 0.2$
w_2	0.01	0.01	0.24	0.2	$40.9 \pm 0.9 \pm 0.2$
w_3	0.05	0.00	0.19	0.2	$28.3 \pm 0.7 \pm 0.2$
w_4	0.04	0.01	0.21	0.2	$15.6 \pm 0.8 \pm 0.2$
w_5	0.03	0.01	0.17	0.2	$11.9 \pm 0.7 \pm 0.2$
w_6	0.02	0.01	0.15	0.2	$5.4 \pm 0.6 \pm 0.2$
w_7	0.00	0.00	0.06	0.1	$1.2 \pm 0.2 \pm 0.1$
μ_1	0.03	0.02	0.46	0.5	$0.2 \pm 1.8 \pm 0.5$

Table 5 continued

Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
μ_2	0.03	0.05	0.49	0.5	$0.3 \pm 1.9 \pm 0.5$
μ_3	0.00	0.02	0.43	0.4	$1.5 \pm 1.6 \pm 0.4$
μ_4	0.10	0.09	0.55	0.6	$-3.0 \pm 2.1 \pm 0.6$
μ_5	0.03	0.06	0.51	0.5	$-3.1 \pm 2.0 \pm 0.5$
μ_6	0.09	0.10	0.56	0.6	$-0.6 \pm 2.2 \pm 0.6$
μ_7	0.08	0.07	0.42	0.4	$-0.6 \pm 1.6 \pm 0.4$
Δw_1	0.10	0.09	0.46	0.5	$7.1 \pm 1.8 \pm 0.5$
Δw_2	0.08	0.07	0.48	0.5	$5.8 \pm 1.9 \pm 0.5$
Δw_3	0.06	0.05	0.39	0.4	$1.3 \pm 1.5 \pm 0.4$
Δw_4	0.04	0.05	0.43	0.4	$-1.2 \pm 1.6 \pm 0.4$
Δw_5	0.06	0.04	0.35	0.4	$2.0 \pm 1.4 \pm 0.4$
Δw_6	0.06	0.05	0.31	0.3	$1.8 \pm 1.2 \pm 0.3$
Δw_7	0.01	0.01	0.12	0.1	$0.5 \pm 0.5 \pm 0.1$

Table 6 Results for ε_i, w_i, μ_i, and Δw_i for the DNN tagger: systematic uncertainties associated with the ΔE parametrization, fit bias, and fit model, and total systematic uncertainty are shown together with the fit results (with stat. and syst. uncertainty). The results are given in percent

Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
$B^0 \to D^{(*)-} h^+$					
ε_1	0.02	0.01	0.09	0.1	$14.3 \pm 0.3 \pm 0.1$
ε_2	0.00	0.01	0.09	0.1	$17.9 \pm 0.3 \pm 0.1$
ε_3	0.01	0.02	0.10	0.1	$22.5 \pm 0.4 \pm 0.1$
ε_4	0.00	0.00	0.07	0.1	$11.0 \pm 0.3 \pm 0.1$
ε_5	0.00	0.00	0.07	0.1	$10.4 \pm 0.3 \pm 0.1$
ε_6	0.01	0.00	0.07	0.1	$9.6 \pm 0.2 \pm 0.1$
ε_7	0.02	0.01	0.08	0.1	$14.2 \pm 0.3 \pm 0.1$
w_1	0.05	0.07	0.51	0.5	$48.2 \pm 1.9 \pm 0.5$
w_2	0.02	0.02	0.46	0.5	$43.6 \pm 1.6 \pm 0.5$
w_3	0.03	0.02	0.40	0.4	$33.9 \pm 1.4 \pm 0.4$
w_4	0.08	0.06	0.54	0.6	$19.3 \pm 1.9 \pm 0.6$
w_5	0.05	0.04	0.52	0.5	$19.7 \pm 1.9 \pm 0.5$
w_6	0.07	0.04	0.49	0.5	$10.8 \pm 1.8 \pm 0.5$
w_7	0.02	0.04	0.36	0.4	$3.5 \pm 1.2 \pm 0.4$
μ_1	0.35	0.24	1.02	1.1	$3.7 \pm 3.7 \pm 1.1$
μ_2	0.12	0.08	0.92	0.9	$7.3 \pm 3.2 \pm 0.9$
μ_3	0.00	0.02	0.81	0.8	$4.6 \pm 2.9 \pm 0.8$
μ_4	0.39	0.35	1.14	1.3	$2.2 \pm 4.0 \pm 1.3$
μ_5	0.12	0.16	1.11	1.1	$7.4 \pm 4.1 \pm 1.1$
μ_6	0.16	0.06	1.10	1.1	$1.5 \pm 4.1 \pm 1.1$
μ_7	0.12	0.07	0.80	0.8	$-2.5 \pm 3.1 \pm 0.8$
Δw_1	0.13	0.11	0.64	0.7	$-1.1 \pm 2.3 \pm 0.7$
Δw_2	0.18	0.16	0.58	0.6	$5.6 \pm 2.1 \pm 0.6$
Δw_3	0.01	0.01	0.53	0.5	$7.1 \pm 1.8 \pm 0.5$
Δw_4	0.08	0.05	0.77	0.8	$4.5 \pm 2.8 \pm 0.8$
Δw_5	0.06	0.03	0.78	0.8	$7.9 \pm 2.7 \pm 0.8$
Δw_6	0.09	0.12	0.80	0.8	$5.7 \pm 3.0 \pm 0.8$
Δw_7	0.12	0.09	0.61	0.6	$3.4 \pm 2.3 \pm 0.6$

Table 6 continued

Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
$B^+ \to \bar{D}^{(*)0} h^+$					
ε_1	0.01	0.01	0.07	0.1	$12.7 \pm 0.3 \pm 0.1$
ε_2	0.04	0.00	0.07	0.1	$16.2 \pm 0.3 \pm 0.1$
ε_3	0.02	0.01	0.08	0.1	$21.2 \pm 0.3 \pm 0.1$
ε_4	0.01	0.00	0.06	0.1	$11.0 \pm 0.2 \pm 0.1$
ε_5	0.01	0.01	0.06	0.1	$11.3 \pm 0.2 \pm 0.1$
ε_6	0.01	0.01	0.06	0.1	$10.3 \pm 0.2 \pm 0.1$
ε_7	0.08	0.01	0.07	0.1	$17.3 \pm 0.2 \pm 0.1$
w_1	0.01	0.00	0.27	0.3	$47.2 \pm 1.1 \pm 0.3$
w_2	0.01	0.00	0.24	0.2	$40.9 \pm 0.9 \pm 0.2$
w_3	0.06	0.02	0.19	0.2	$29.8 \pm 0.7 \pm 0.2$
w_4	0.02	0.02	0.23	0.2	$17.2 \pm 0.9 \pm 0.2$
w_5	0.06	0.02	0.19	0.2	$11.9 \pm 0.7 \pm 0.2$
w_6	0.01	0.02	0.14	0.1	$4.6 \pm 0.5 \pm 0.1$
w_7	0.01	0.00	0.06	0.1	$1.8 \pm 0.2 \pm 0.1$
μ_1	0.08	0.10	0.54	0.6	$0.4 \pm 2.1 \pm 0.6$
μ_2	0.11	0.08	0.49	0.5	$3.2 \pm 1.9 \pm 0.5$
μ_3	0.04	0.02	0.42	0.4	$4.4 \pm 1.6 \pm 0.4$
μ_4	0.01	0.04	0.57	0.6	$-1.4 \pm 2.2 \pm 0.6$
μ_5	0.02	0.02	0.54	0.5	$-4.1 \pm 2.1 \pm 0.5$
μ_6	0.07	0.05	0.53	0.5	$-5.4 \pm 2.1 \pm 0.5$
μ_7	0.04	0.03	0.37	0.4	$-4.5 \pm 1.5 \pm 0.4$
Δw_1	0.02	0.04	0.54	0.5	$0.8 \pm 2.1 \pm 0.5$
Δw_2	0.00	0.03	0.48	0.5	$6.4 \pm 1.8 \pm 0.5$
Δw_3	0.05	0.08	0.39	0.4	$5.3 \pm 1.5 \pm 0.4$
Δw_4	0.11	0.12	0.46	0.5	$3.4 \pm 1.8 \pm 0.5$
Δw_5	0.14	0.11	0.38	0.4	$2.5 \pm 1.5 \pm 0.4$
Δw_6	0.05	0.03	0.29	0.3	$0.5 \pm 1.1 \pm 0.3$
Δw_7	0.01	0.02	0.11	0.1	$0.4 \pm 0.5 \pm 0.1$

they do not cause statistically significant bias for samples up to 700 fb^{-1} by performing fits to generic simulation.

11 Linearity check

We check whether the dilution r provided by the flavor tagger corresponds to the actual definition $r := 1-2w$ by performing a linearity check. Figure 13 shows the linearity check for both flavor taggers in simulation and data.

For simulation, we determine the true wrong-tag fraction $w_{\rm MC}$ by comparing the MC truth with the flavor-tagger output, and calculate the true dilution $r_{\rm MC} = 1 - 2w_{\rm MC}$. The mean dilution $\langle r_{\rm FBDT}\rangle$ is simply the mean of $|q \cdot r_{\rm FBDT}|$ for correctly associated MC events in each r bin. For data, we obtain the mean $\langle r_{\rm FBDT}\rangle = \langle |q \cdot r_{\rm FBDT}|\rangle$ values from the signal $q \cdot r_{\rm FBDT}$ distribution provided by the $s\mathcal{P}lot$ analysis in Sect. 9. The dilution $r = 1-2\cdot w$ in data is obtained from the fit results for w. The linearity verifies the equivalence on average between the dilution provided by the flavor tagger and the measured one within the uncertainties. For charged B candidates, we observe a slightly non-linear behavior, which is attributed to the fact that both flavor taggers are optimized and trained only for neutral B mesons. However, we observe a good agreement between data and simulation for both neutral and charged B candidates.

12 Comparison with the previous Belle algorithm

A comparison of the current results with the latest results on flavor tagging obtained by Belle [33] provides interesting insight about the current and projected performance of Belle II. We compare partial efficiencies, wrong-tag fractions, total effective efficiencies, and wrong-tag asymmetries in each r-bin. Table 9, and Figs. 14 and 15 compare

Table 7 Effective efficiencies for the category-based (FBDT) tagger: systematic uncertainties associated with the ΔE parametrization, fit bias, and fit model, and total systematic uncertainty are shown together with the fit results (with stat. and syst. uncertainty). The results are given in percent

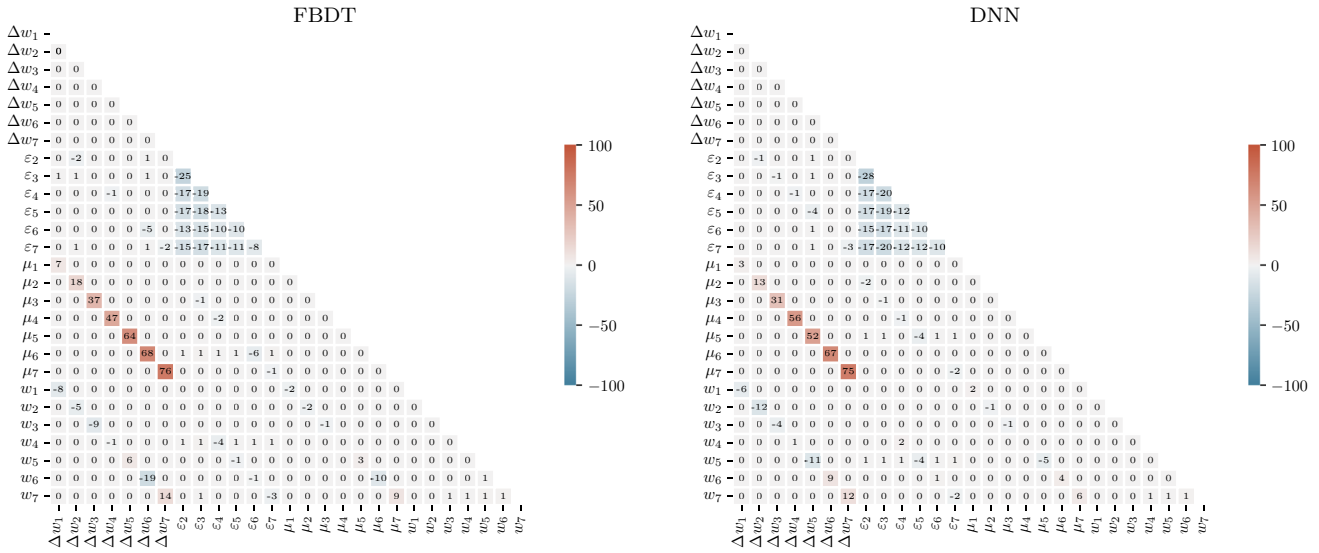
Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
$B^0 \to D^{(*)-}h^+$					
$\varepsilon_{\text{eff},1}$	0.02	0.02	0.02	0.0	$0.1 \pm 0.1 \pm 0.0$
$\varepsilon_{\text{eff},2}$	0.02	0.02	0.06	0.1	$0.5 \pm 0.2 \pm 0.1$
$\varepsilon_{\text{eff},3}$	0.03	0.01	0.13	0.1	$3.3 \pm 0.5 \pm 0.1$
$\varepsilon_{\text{eff},4}$	0.04	0.04	0.14	0.2	$3.3 \pm 0.5 \pm 0.2$
$\varepsilon_{\text{eff},5}$	0.01	0.00	0.16	0.2	$6.1 \pm 0.6 \pm 0.2$
$\varepsilon_{\text{eff},6}$	0.00	0.00	0.16	0.2	$5.4 \pm 0.5 \pm 0.2$
$\varepsilon_{\text{eff},7}$	0.03	0.01	0.20	0.2	$11.3 \pm 0.6 \pm 0.2$
Total	0.06	0.05	0.36	0.4	$30.0 \pm 1.2 \pm 0.4$
$B^+ \to \bar{D}^{(*)0}h^+$					
$\varepsilon_{\text{eff},1}$	0.00	0.00	0.01	0.0	$0.0 \pm 0.0 \pm 0.0$
$\varepsilon_{\text{eff},2}$	0.00	0.00	0.03	0.0	$0.5 \pm 0.1 \pm 0.0$
$\varepsilon_{\text{eff},3}$	0.01	0.00	0.07	0.1	$3.8 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},4}$	0.01	0.00	0.07	0.1	$5.4 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},5}$	0.01	0.01	0.08	0.1	$7.0 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},6}$	0.02	0.00	0.06	0.1	$7.1 \pm 0.2 \pm 0.1$
$\varepsilon_{\text{eff},7}$	0.07	0.01	0.06	0.1	$13.2 \pm 0.2 \pm 0.1$
Total	0.08	0.02	0.16	0.2	$37.0 \pm 0.6 \pm 0.2$

Table 8 Effective efficiencies for the DNN tagger: systematic uncertainties associated with the ΔE parametrization, fit bias, and fit model, and total systematic uncertainty are shown together with the fit results (with stat. and syst. uncertainty). The results are given in percent

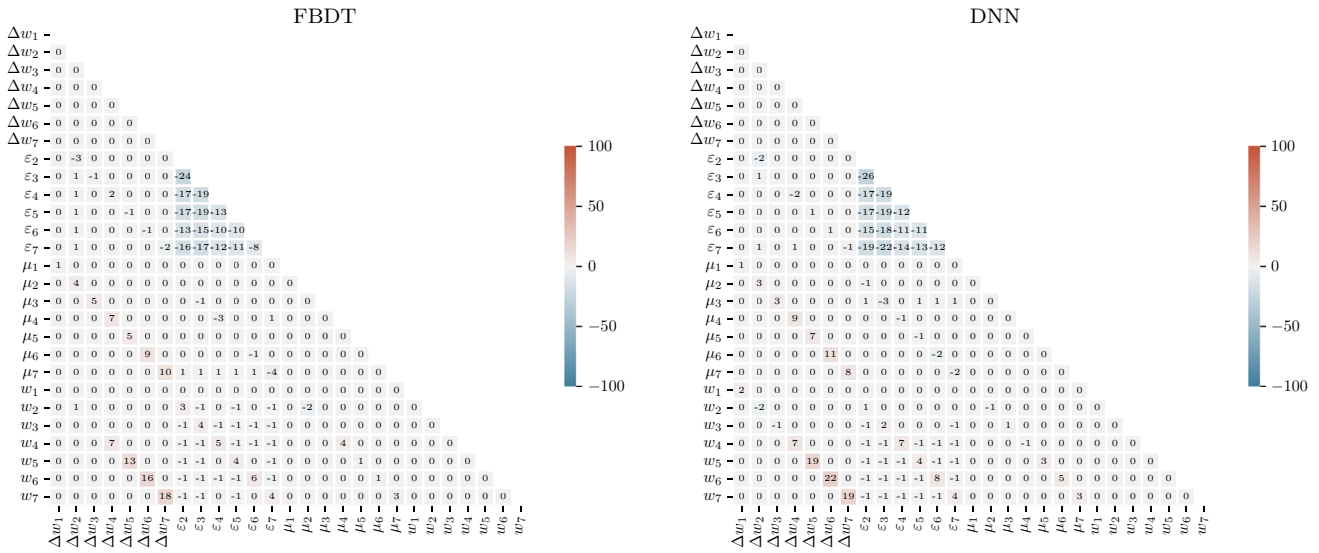
Parameter	ΔE PDF	Fit bias	Model	Total syst. uncty.	Fit results
$B^0 \to D^{(*)-}h^+$					
$\varepsilon_{\text{eff},1}$	0.02	0.02	0.01	< 0.1	$0.0 \pm 0.0 \pm 0.0$
$\varepsilon_{\text{eff},2}$	0.01	0.01	0.05	0.1	$0.3 \pm 0.1 \pm 0.1$
$\varepsilon_{\text{eff},3}$	0.00	0.00	0.13	0.1	$2.3 \pm 0.4 \pm 0.1$
$\varepsilon_{\text{eff},4}$	0.00	0.00	0.12	0.1	$4.2 \pm 0.5 \pm 0.1$
$\varepsilon_{\text{eff},5}$	0.01	0.01	0.15	0.2	$3.8 \pm 0.5 \pm 0.2$
$\varepsilon_{\text{eff},6}$	0.02	0.02	0.16	0.2	$5.9 \pm 0.6 \pm 0.2$
$\varepsilon_{\text{eff},7}$	0.03	0.01	0.23	0.2	$12.3 \pm 0.7 \pm 0.2$
Total	0.05	0.03	0.37	0.4	$28.8 \pm 1.2 \pm 0.4$
$B^+ \to \bar{D}^{(*)0}h^+$					
$\varepsilon_{\text{eff},1}$	0.00	0.00	0.01	< 0.1	$0.0 \pm 0.0 \pm 0.0$
$\varepsilon_{\text{eff},2}$	0.00	0.00	0.03	< 0.1	$0.5 \pm 0.1 \pm 0.0$
$\varepsilon_{\text{eff},3}$	0.02	0.01	0.07	0.1	$3.5 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},4}$	0.01	0.00	0.07	0.1	$4.7 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},5}$	0.01	0.00	0.07	0.1	$6.6 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},6}$	0.01	0.01	0.07	0.1	$8.5 \pm 0.3 \pm 0.1$
$\varepsilon_{\text{eff},7}$	0.09	0.00	0.07	0.1	$16.1 \pm 0.3 \pm 0.1$
Total	0.09	0.01	0.16	0.2	$39.9 \pm 0.6 \pm 0.2$

Belle II and Belle results. The Belle flavor tagger, which was a category-based algorithm, reached a total effective efficiency of $(30.1 \pm 0.4)\%$ on Belle data [33]. We observe about the same or slightly better performance than Belle in all bins except in the highest r bin, for which we observe a smaller partial efficiency and also a slightly worse performance with respect to expectations (see Figs. 7 and 8). For the wrong-tag fractions, we observe larger asymmetries between B^0 and $\bar{B}^0$ than Belle.

In comparison with the Belle algorithm [32,33], the Belle II category-based flavor tagger considers more flavor signatures and more input variables, and is fully based on multivariate methods avoiding the cut-based identification



(a) $B^0 \to D^{(*)-}h^+$.



(b) $B^+ \to \overline{D}^{(*)0}h^+$.

Fig. 12 Correlation coefficients between ε_i, Δw_i, Δw_i, and μ_i for the (left) category-based and (right) DNN flavor tagger in data. The results are shown for (top) neutral and (bottom) charged $B \to D^{(*)}h^+$ candidates

of decay products and exploiting the correlations between input variables and between flavor signatures.

For the current version of the category-based and the DNN flavor taggers, we obtain a total effective efficiency around 32.5% in simulation. Previous studies [1,31,34] show that using track impact parameters as additional input variables potentially improves the total effective tagging efficiency by about 2 to 3% in its absolute value. Nonetheless, the current results show that Belle II can reach a tagging performance comparable with the one obtained by Belle even with a not yet fully-optimized calibration of the tracking and PID systems, and operating in harsher background conditions than those experienced by Belle.

13 Summary

We report on the performance of two new Belle II B-flavor tagging algorithms on Belle II data collected at the $\Upsilon(4S)$ resonance between 2019 and 2020. The algorithms exploit

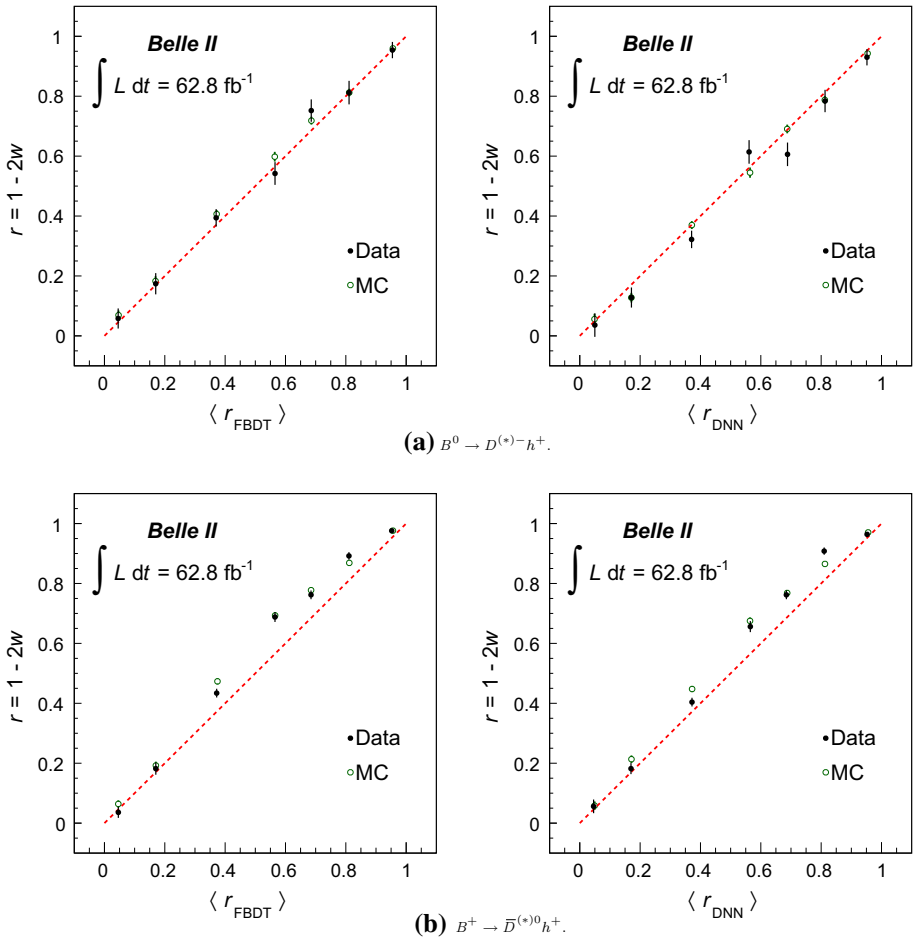


Fig. 13 Dilution factor $r = 1 - 2w$ as a function of the mean dilution $\langle |q \cdot r| \rangle$ provided by the (left) category-based and (right) DNN flavor tagger in data and MC simulation for (top) neutral and (bottom) charged $B \to D^{(*)} h^+$ candidates. The red dashed lines correspond to a linear function with an intercept at zero and a slope of one, corresponding to a perfect agreement between predicted and measured dilution

modern machine-learning techniques to determine the quark-flavor content of neutral B mesons from the kinematic, track-hit, and particle-identification information associated with the reconstructed decay products. We validate the algorithms in simulation and in data using samples containing one fully reconstructed signal B decay. We reconstruct abundant signal B decays to flavor-specific hadronic final states and then use the remaining tracks and neutral clusters in each event as input for the flavor taggers.

We use the ΔE distribution of the fully reconstructed B candidates, restricted in $M_{\rm bc}$, to identify the B signals and measure the tagging efficiencies, fractions of wrongly tagged events and related asymmetries from the flavor evolution of the signal $B\bar{B}$ pairs in a time-integrated way. Using a category-based flavor tagging algorithm, we obtain for neutral B candidates the total effective efficiency

$$\varepsilon_{\rm eff} = \big(30.0 \pm 1.2(\text{stat}) \pm 0.4(\text{syst})\big)\%,$$

and for charged B candidates

$$\varepsilon_{\rm eff} = \big(37.0 \pm 0.6(\text{stat}) \pm 0.2(\text{syst})\big)\%.$$

Using a deep-learning-based flavor tagging algorithm, we obtain for neutral B candidates the total effective efficiency

$$\varepsilon_{\rm eff} = \big(28.8 \pm 1.2(\text{stat}) \pm 0.4(\text{syst})\big)\%,$$

and for charged B candidates

$$\varepsilon_{\rm eff} = \big(39.9 \pm 0.6(\text{stat}) \pm 0.2(\text{syst})\big)\%.$$

The performance of the flavor taggers is generally compatible with expectations from simulation and is comparable with the best performance obtained by the Belle experiment within the uncertainties. While both flavor taggers perform equally good in simulation, the deep-learning-based algorithm performs slightly worse than the category-based one in data. This is most likely due to current discrepancies between data and simulation since deep-learning methods heavily rely on a good description of the dependences among input variables. Thus we expect improvements in the future.

This work marks a milestone for future calibrations, which will play an essential role in measurements of CP asymmetries at Belle II and ultimately in the search for deviations from the Standard Model expectations.

Table 9 Partial efficiencies ε_i, wrong-tag fractions w_i, total effective efficiencies $\varepsilon_{\text{eff},i}$, tagging efficiency asymmetries $\mu = \Delta\varepsilon/(2\varepsilon)$, and wrong-tag fraction asymmetries Δw obtained with the Belle II category-based (FBDT) and deep-learning (DNN) flavor taggers in 2019–2020 Belle II data and with the Belle flavor tagger in 2003–2010 Belle data [33] taken with the second silicon-vertex detector configuration (SVD2). There are no available Belle results for μ. Statistical and systematical uncertainties are added in quadrature. All values are given in percent

	$\varepsilon_i \pm \delta\varepsilon_i$			$w_i \pm \delta w_i$		
r- Interval	FBDT	DNN	Belle	FBDT	DNN	Belle
0.000–0.100	19.0 ± 0.3	14.3 ± 0.3	22.2 ± 0.4	47.1 ± 1.7	48.2 ± 2.0	50.0
0.100–0.250	17.1 ± 0.3	17.9 ± 0.3	14.5 ± 0.3	41.3 ± 1.8	43.6 ± 1.7	41.9 ± 0.4
0.250–0.500	21.3 ± 0.3	22.5 ± 0.4	17.7 ± 0.4	30.3 ± 1.5	33.9 ± 1.5	31.9 ± 0.3
0.500–0.625	11.3 ± 0.3	11.0 ± 0.3	11.5 ± 0.3	22.9 ± 2.0	19.3 ± 2.0	22.3 ± 0.4
0.625–0.750	10.7 ± 0.3	10.4 ± 0.3	10.2 ± 0.3	12.4 ± 1.9	19.7 ± 2.0	16.3 ± 0.4
0.750–0.875	8.2 ± 0.2	9.6 ± 0.2	8.7 ± 0.3	9.4 ± 2.0	10.8 ± 1.9	10.4 ± 0.4
0.875–1.000	12.4 ± 0.2	14.2 ± 0.3	15.3 ± 0.3	2.3 ± 1.4	3.5 ± 1.4	2.5 ± 0.3

$\varepsilon_{\text{eff},i} \pm \delta\varepsilon_{\text{eff},i}$			
r-Interval	FBDT	DNN	Belle
0.000–0.100	0.1 ± 0.1	0.0 ± 0.1	0.0
0.100–0.250	0.5 ± 0.2	0.3 ± 0.1	0.4 ± 0.1
0.250–0.500	3.3 ± 0.5	2.3 ± 0.4	2.3 ± 0.1
0.500–0.625	3.3 ± 0.5	4.2 ± 0.5	3.5 ± 0.1
0.625–0.750	6.1 ± 0.6	3.8 ± 0.5	4.6 ± 0.2
0.750–0.875	5.4 ± 0.5	5.9 ± 0.6	5.5 ± 0.1
0.875–1.000	11.3 ± 0.6	12.3 ± 0.7	13.8 ± 0.3
Total	30.0 ± 1.3	28.8 ± 1.3	30.1 ± 0.4

	$\mu_i \pm \delta\mu_i$			$\Delta w_i \pm \delta\Delta w_i$		
r- Interval	FBDT	DNN	Belle	FBDT	DNN	Belle
0.000–0.100	4.4 ± 3.4	3.7 ± 3.9	–	8.8 ± 2.1	-1.1 ± 2.4	0.0
0.100–0.250	3.9 ± 3.4	7.3 ± 3.4	–	6.1 ± 2.2	5.6 ± 2.2	-0.9 ± 0.4
0.250–0.500	6.8 ± 3.0	4.6 ± 3.0	–	2.7 ± 2.0	7.1 ± 2.0	1.0 ± 0.4
0.500–0.625	3.2 ± 4.1	2.2 ± 4.2	–	5.5 ± 2.8	4.5 ± 2.9	-1.1 ± 0.4
0.625–0.750	-0.5 ± 4.2	7.4 ± 4.2	–	0.7 ± 3.0	7.9 ± 2.9	-1.9 ± 0.5
0.750–0.875	10.8 ± 4.4	1.5 ± 4.2	–	7.7 ± 3.3	5.7 ± 3.1	1.7 ± 0.4
0.875–1.000	-3.7 ± 3.4	-2.5 ± 4.2	–	0.6 ± 2.5	3.4 ± 2.5	-0.4 ± 0.2

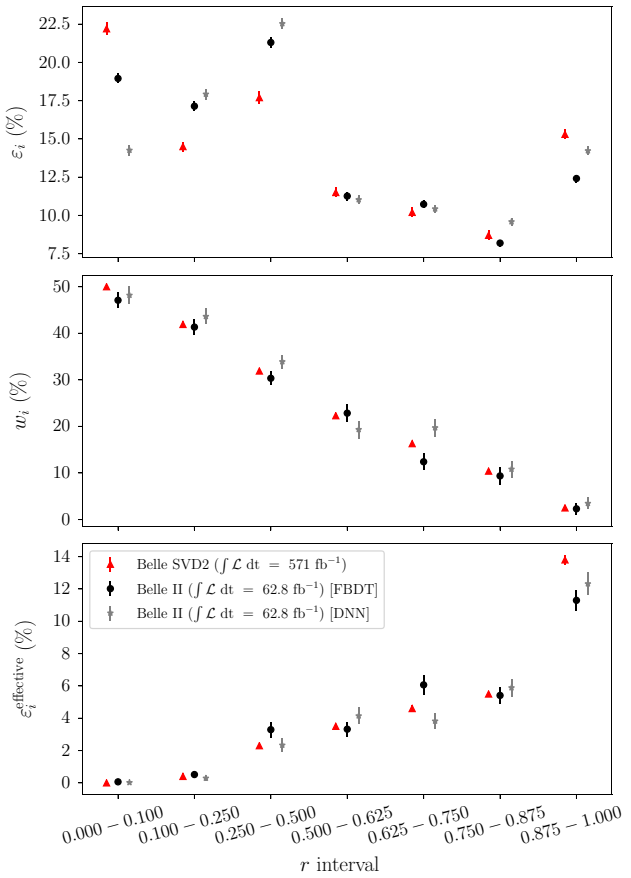


Fig. 14 Performance of the Belle II category-based (FBDT) and deep-learning (DNN) flavor taggers in 2019–2020 Belle II data and of the Belle flavor tagger in 2003–2010 Belle data [33] taken with the second silicon-vertex detector configuration (SVD2)

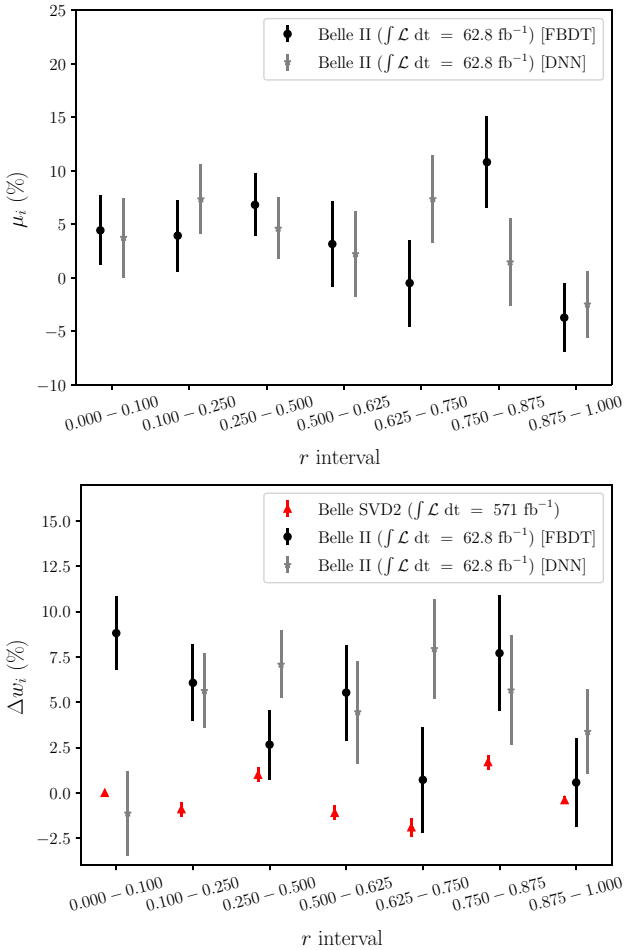


Fig. 15 Comparison of (top) tagging efficiency asymmetries μ_i and (bottom) wrong-tag fraction asymmetries Δw_i for the Belle II category-based (FBDT) and deep-learning (DNN) flavor taggers in 2019–2020 Belle II data and for the Belle flavor tagger in 2003–2010 Belle data [33] taken with the second silicon-vertex detector configuration (SVD2). There are no available Belle results for μ

Acknowledgements We thank the SuperKEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; the KEK computer group for on-site computing support; and the raw-data centers at BNL, DESY, GridKa, IN2P3, and INFN for off-site computing support. This work was supported by the following funding sources: Science Committee of the Republic of Armenia Grant No. 20TTCG-1C010; Australian Research Council and research grant Nos. DP180102629, DP170102389, DP170102204, DP150103061, FT130100303, FT130100018, and FT120100745; Austrian Federal Ministry of Education, Science and Research, Austrian Science Fund No. P 31361-N36, and Horizon 2020 ERC Starting Grant no. 947006 "InterLeptons"; Natural Sciences and Engineering Research Council of Canada, Compute Canada and CANARIE; Chinese Academy of Sciences and research grant No. QYZDJ-SSW-SLH011, National Natural Science Foundation of China and research grant Nos. 11521505, 11575017, 11675166, 11761141009, 11705209, and 11975076, LiaoNing Revitalization Talents Program under contract No. XLYC1807135, Shanghai Municipal Science and Technology Committee under contract No. 19ZR1403000, Shanghai Pujiang Program under Grant No. 18PJ1401000, and the CAS Center for Excellence in Particle Physics (CCEPP); the Ministry of Education, Youth and Sports of the Czech Republic under Contract No. LTT17020 and Charles University grants SVV 260448 and GAUK 404316; European Research Council, 7th Framework PIEF-GA-2013-622527, Horizon 2020 ERC Advanced Grants No. 267104 and 884719, Horizon 2020 ERC Consolidator Grant No. 819127, Horizon 2020 Marie Sklodowska-Curie grant agreement No. 700525 'NIOBE,' and Horizon 2020 Marie Sklodowska-Curie RISE project JENNIFER2 grant agreement No. 822070 (European grants); L'Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) du CNRS (France); BMBF, DFG, HGF, MPG, and AvH Foundation (Germany); Department of Atomic Energy under Project Identification No. RTI 4002 and Department of Science and Technology (India); Israel Science Foundation grant No. 2476/17, United States-Israel Binational Science Foundation grant No. 2016113, and Israel Ministry of Science grant No. 3-16543; Istituto Nazionale di Fisica Nucleare and the research grants BELLE2; Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research grant Nos. 16H03968, 16H03993, 16H06492, 16K05323, 17H01133, 17H05405, 18K03621, 18H03710, 18H05226, 19H00682, 26220706, and 26400255, the National Institute of Informatics, and Science Information NETwork 5 (SINET5), and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan; National Research Foundation (NRF) of Korea Grant Nos. 2016R1D1A1B01010135, 2016R1D1A1B02012900, 2018R1A2B3003643, 2018R1A6A1A06024970, 2018R1D1A1B07047294, 2019K1A3A7A09033840, and 2019R1I1A3A01058933, Radiation Science Research Institute, For-

eign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD; Universiti Malaya RU grant, Akademi Sains Malaysia and Ministry of Education Malaysia; Frontiers of Science Program contracts FOINS-296, CB-221329, CB-236394, CB-254409, and CB-180023, and SEP-CINVESTAV research grant 237 (Mexico); the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation, Agreement 14.W03.31.0026, and the HSE University Basic Research Program, Moscow; University of Tabuk research grants S-0256-1438 and S-0280-1439 (Saudi Arabia); Slovenian Research Agency and research grant Nos. J1-9124 and P1-0135; Agencia Estatal de Investigacion, Spain grant Nos. FPA2014-55613-P and FPA2017-84445-P, and CIDEGENT/2018/020 of Generalitat Valenciana; Ministry of Science and Technology and research grant Nos. MOST106-2112-M-002-005-MY3 and MOST107-2119-M-002-035-MY3, and the Ministry of Education (Taiwan); Thailand Center of Excellence in Physics; TUBITAK ULAKBIM (Turkey); Ministry of Education and Science of Ukraine; the US National Science Foundation and research grant Nos. PHY-1807007 and PHY-1913789, and the US Department of Energy and research grant Nos. DE-AC06-76RLO1830, DE-SC0007983, DE-SC0009824, DE-SC0009973, DE-SC0010007, DE-SC0010073, DE-SC0010118, DE-SC0010504, DE-SC0011784, DE-SC0012704, DE-SC0021274; and the Vietnam Academy of Science and Technology (VAST) under grant DL0000.05/21-23.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: Belle II data is currently accessible only for members of the Belle II collaboration.]

Funded by SCOAP[3].

References

1. E. Kou et al., PTEP **2019**(12), 123C01 (2019). https://doi.org/10.1093/ptep/ptz106 [Erratum: PTEP 2020, 029201 (2020)]
2. K. Abe et al., Phys. Rev. D **71**, 072003 (2005). https://doi.org/10.1103/PhysRevD.71.072003 [Erratum: Phys. Rev. D 71, 079903 (2005)]
3. B. Aubert et al., Phys. Rev. D **67**, 072002 (2003). https://doi.org/10.1103/PhysRevD.67.072002
4. R. Aaij et al., Eur. Phys. J. C **76**(7), 412 (2016). https://doi.org/10.1140/epjc/s10052-016-4250-2
5. I. Adachi et al., Phys. Rev. Lett. **108**, 171802 (2012). https://doi.org/10.1103/PhysRevLett.108.171802
6. B. Aubert et al., Phys. Rev. D **79**, 072009 (2009). https://doi.org/10.1103/PhysRevD.79.072009
7. R. Aaij et al., Phys. Rev. Lett. **115**(3), 031601 (2015). https://doi.org/10.1103/PhysRevLett.115.031601
8. B. Aubert et al., Phys. Rev. D **71**, 032005 (2005). https://doi.org/10.1103/PhysRevD.71.032005
9. I. Adachi et al., Phys. Rev. D **98**(11), 112012 (2018). https://doi.org/10.1103/PhysRevD.98.112012
10. I. Adachi et al., Phys. Rev. D **88**(9), 092003 (2013). https://doi.org/10.1103/PhysRevD.88.092003
11. J.P. Lees et al., Phys. Rev. D **87**(5), 052009 (2013). https://doi.org/10.1103/PhysRevD.87.052009
12. P. Vanhoefer et al., Phys. Rev. D **93**(3), 032010 (2016). https://doi.org/10.1103/PhysRevD.94.099903, https://doi.org/10.1103/PhysRevD.93.032010 [Addendum: Phys. Rev.D 94(9), 099903 (2016)]
13. B. Aubert et al., Phys. Rev. D **76**, 052007 (2007). https://doi.org/10.1103/PhysRevD.76.052007
14. A. Kusaka et al., Phys. Rev. Lett. **98**, 221602 (2007). https://doi.org/10.1103/PhysRevLett.98.221602
15. J.P. Lees et al., Phys. Rev. D **88**(1), 012003 (2013). https://doi.org/10.1103/PhysRevD.88.012003
16. K.F. Chen et al., Conf. Proc. C **060726**, 823 (2006). https://doi.org/10.1103/PhysRevLett.98.031802
17. J.P. Lees et al., Phys. Rev. D **85**, 112010 (2012). https://doi.org/10.1103/PhysRevD.85.112010
18. M. Fujikawa et al., Phys. Rev. D **81**, 011101 (2010). https://doi.org/10.1103/PhysRevD.81.011101
19. B. Aubert et al., Phys. Rev. D **79**, 052003 (2009). https://doi.org/10.1103/PhysRevD.79.052003
20. B. Aubert et al., Phys. Rev. D **78**, 071102 (2008). https://doi.org/10.1103/PhysRevD.78.071102
21. Y. Ushiroda et al., Phys. Rev. D **74**, 111104 (2006). https://doi.org/10.1103/PhysRevD.74.111104
22. N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963). https://doi.org/10.1103/PhysRevLett.10.531
23. M. Kobayashi, T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973). https://doi.org/10.1143/PTP.49.652
24. A.B. Carter, A.I. Sanda, Phys. Rev. D **23**, 1567 (1981). https://doi.org/10.1103/PhysRevD.23.1567
25. C. Dib, I. Dunietz, F.J. Gilman, Y. Nir, Phys. Rev. D **41**, 1522 (1990). https://doi.org/10.1103/PhysRevD.41.1522
26. M. Gronau, D. London, Phys. Rev. Lett. **65**, 3381 (1990). https://doi.org/10.1103/PhysRevLett.65.3381
27. H.J. Lipkin, Y. Nir, H.R. Quinn, A. Snyder, Phys. Rev. D **44**, 1454 (1991). https://doi.org/10.1103/PhysRevD.44.1454
28. D. Atwood, M. Gronau, A. Soni, Phys. Rev. Lett. **79**, 185 (1997). https://doi.org/10.1103/PhysRevLett.79.185
29. M. Gronau, Phys. Lett. B **627**, 82 (2005). https://doi.org/10.1016/j.physletb.2005.09.014
30. M. Beneke, Phys. Lett. B **620**, 143 (2005). https://doi.org/10.1016/j.physletb.2005.06.045
31. F. Abudinén, Development of a B^o flavor tagger and performance study of a novel time-dependent CP analysis of the decay $B^o \to \pi^0\pi^0$ at Belle II. Ph.D. thesis, Ludwig-Maximilians-Universität München (2018). https://docs.belle2.org/record/1215/files/BELLE2-PTHESIS-2018-003.pdf
32. H. Kakuno et al., Nucl. Instrum. Meth. A **533**, 516 (2004). https://doi.org/10.1016/j.nima.2004.06.159
33. A.J. Bevan et al., Eur. Phys. J. C **74**, 3026 (2014). https://doi.org/10.1140/epjc/s10052-014-3026-9
34. J.F. Gemmler, Development and Deployment of a Deep Neural Network based Flavor Tagger for Belle II. Ph.D. thesis, Karlsruher Institut für Technologie (2020). https://docs.belle2.org/record/2423/files/BELLE2-PTHESIS-2021-005.pdf
35. T. Abe et al., Belle II Technical Design Report (2010)
36. K. Akai, K. Furukawa, H. Koiso, Nucl. Instrum. Meth. A **907**, 188 (2018). https://doi.org/10.1016/j.nima.2018.08.017

37. T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, N. Braun, Comput. Softw. Big Sci. **3**(1), 1 (2019). https://doi.org/10.1007/s41781-018-0017-9
38. A.T. Ryd et al., EvtGen: A Monte Carlo Generator for B-Physics (2005)
39. S. Agostinelli et al., Nucl. Instrum. Meth. A **506**, 250 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8
40. P. Zyla et al., PTEP **2020**(8), 083C01 (2020). https://doi.org/10.1093/ptep/ptaa104
41. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, Comput. Phys. Commun. **191**, 159 (2015). https://doi.org/10.1016/j.cpc.2015.01.024
42. P.M. Lewis et al., Nucl. Instrum. Meth. A **914**, 69 (2019). https://doi.org/10.1016/j.nima.2018.05.071
43. R. Kleiss, H. Burkhardt, Comput. Phys. Commun. **81**, 372 (1994). https://doi.org/10.1016/0010-4655(94)90085-X
44. C.M. Carloni Calame, C. Lunardini, G. Montagna, O. Nicrosini, F. Piccinini, Nucl. Phys. B **584**, 459 (2000). https://doi.org/10.1016/S0550-3213(00)00356-4
45. F.A. Berends, P.H. Daverveldt, R. Kleiss, Nucl. Phys. B **253**, 441 (1985). https://doi.org/10.1016/0550-3213(85)90541-3
46. F. Abudinén et al., Chin. Phys. C **44**(2), 021001 (2020). https://doi.org/10.1088/1674-1137/44/2/021001
47. V. Bertacchi et al., Comput. Phys. Commun. **259**, 107610 (2021). https://doi.org/10.1016/j.cpc.2020.107610
48. T. Keck, Comput. Softw. Big Sci. **1**(1), 2 (2017). https://doi.org/10.1007/s41781-017-0002-8
49. T. Keck, Machine learning algorithms for the Belle II experiment and their validation on Belle data. Ph.D. thesis, KIT, Karlsruhe (2017). https://doi.org/10.1007/978-3-319-98249-6
50. T. Keck et al., Comput. Softw. Big Sci. **3**(1), 6 (2019). https://doi.org/10.1007/s41781-019-0021-8
51. R. Zadeh, B. Ramsundar, *TensorFlow for Deep Learning: From Linear Regression to Reinforcement Learning* (O'Reilly Media, 2018)
52. P. Baldi, P. Sadowski, D. Whiteson, Nat. Commun. **5**, 4308 (2014). https://doi.org/10.1038/ncomms5308
53. D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature **323**(6088), 533 (1986). https://doi.org/10.1038/323533a0
54. Nvidia corporation. GEFORCE GTX 970. https://www.nvidia.com/content/geforce-gtx/GTX_970_User_Guide.pdf
55. J.F. Krohn et al., Nucl. Instrum. Meth. A **976**, 164269 (2020). https://doi.org/10.1016/j.nima.2020.164269
56. G.C. Fox, S. Wolfram, Nucl. Phys. B **149**, 413 (1979). https://doi.org/10.1016/0550-3213(79)90120-2 [Erratum: Nucl. Phys. B 157, 543 (1979)]
57. Y. Amhis et al., Eur. Phys. J. C **77**(12), 895 (2017). https://doi.org/10.1140/epjc/s10052-017-5058-4. Updated results and plots available at https://hflav.web.cern.ch
58. T. Skwarnicki, A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances. Ph.D. thesis, Cracow, INP (1986)
59. M. Pivk, F.R. Le Diberder, Nucl. Instrum. Meth. A **555**, 356 (2005). https://doi.org/10.1016/j.nima.2005.08.106
60. B. Efron, Ann. Stat. **7**(1), 1 (1979). https://doi.org/10.1214/aos/1176344552