Machine-Learning-Based Model for Indicators of the Resource-Based Security of Interests in High-Level Organizational Systems*

Sergiy Skybyk[1,*], Anatoliy Doroshenko[2], Olena Ilyina[1], Igor Sinitsyn[1]

[1]Institute of Software Systems NAS of Ukraine, Glushkov 40, 03187 Kyiv, Ukraine
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi 37, 03056 Kyiv, Ukraine

Abstract

Reliable situational awareness of resource sufficiency is vital for macro-level organizational systems that face deep uncertainty, multi-level decision chains and tight resource constraints. We present an end-to-end framework for a Composite Multi-functional System of Resource-based Security Indicators (RSI) that supports strategic and operational decisions in Multidimensional Management (MDM); the defense domain was considered as its notable representative. At the conceptual layer we formalize the management domain by a two-level network of decision centers governed by Reference Concepts of Management (RCM). Each RCM represents either a strategic goal, a strategic capability, a crisis management task, or an activity ledger. For each RCM we derive individual indicators of performance well-being and resource well-being. Predicate relations between RCM types enable both direct and indirect (integrated or statistically inferred) estimates. Four functional indicators: express resource-risk diagnosis, vulnerability ranking, external-impact criticality and strategic-adaptation diagnosis — are produced by tailored aggregations of individual indicators and context data. We detail the machine-learning pipeline for the first, time-critical function. A heterogeneous ensemble of five classifiers (Bayes, SVM, Random Forest, kNN, Logistic Regression) is trained on historic ledgers and expert audits. To account for the NFL theorem, each classifier is weighted by a model-conformity coefficient (reflects data assumption violations) and by a quality index robust to class imbalance. Forecasting supports a single priority model or a weighted-voting ensemble. Future work will deliver prototypes, quantitative validation, N-ary classification and hybrid time-series forecasting, extending RSI support for high-stakes organizational decisions.

Keywords

Machine-learning classification, Resource-based security metrics, weighted voting ensemble, trategic decision support, KPI monitoring

1. Problem Statement

An integrated presentation of information on the current state of objects and processes — tailored to the needs of decision-makers — is a key element of digital governance. At higher managerial tiers within large-scale Organizational systems (OS) and, especially, national and international projects, this requirement is commonly satisfied by indicator frameworks.

Numerous, largely complementary definitions of an *indicator* [1–3] describe it as a quantitative measure derived from a series of observed facts that characterizes the position of an object within a given domain, including its state of sufficiency or deficiency. Indicators are employed across many domains and serve diverse analytical purposes. For example, the widely used set of international indices of countries' economic and political development [4] aggregates a large number of regularly documented variables, enabling progress tracking, identification of leverage points, and early diagnostics or forecasting.

Indicators for complex systems are typically composite [1, 5, 6]. Their constituent indicators often form a hierarchy in which successive levels refine specific properties, while the composite value is obtained by aggregating constituent scores with weighting coefficients [6]. In high-risk Problem areas

* *Applied Information Systems and Technologies in the Digital Society (AISTDS-2025), October 01, 2025, Kyiv, Ukraine*

[1] * Corresponding author.
sskybyk@gmail.com (S. Skybyk); doroshenkoanatoliy2@gmail.com (A. Doroshenko); ilyina.elena1@ukr.net (O. Ilyina); ips2014@ukr.net (I. Sinitsyn)
0009-0008-4336-680X (S. Skybyk); 0000-0002-8435-1451 (A. Doroshenko); 0000-0002-4073-9649 (O. Ilyina); 0000-0002-4120-0784 (I. Sinitsyn)

(PAs), indicator structures become even more intricate [5, 7]; within the PA of national security, for instance, a composite indicator has been proposed that measures the system's response to deviations from a homeostatic state once critical thresholds of individual components are exceeded [8].

Experience in social-welfare analysis shows the advantages of a composite-indicator system that combines not only various sub-domains but also metrics that differ in their management roles: *social context* (conditions), *social status* (desired outcomes), and *social responsibility* (interventions). Indicators of social well-being likewise integrate characteristics reflecting distinct sufficiency perspectives embedded in alternative definitions of quality of life [9].

A prominent subclass of indicators in management is Key Performance Indicators (KPIs). KPIs are commonly categorized into:

- high-level and low-level KPIs, linked hierarchically, and
- lagging (performance-outcome) and leading (operational) KPIs, where the latter guide actions aimed at improving the former [10].

In decision support, an indicator serves a wide range of analytical needs. Typical examples include comparing different objects — or the same object at different points in time — identifying bottlenecks in development processes, and diagnosing whether the current state of the system is acceptable or satisfactory.

The current indicator value becomes a decision criterion when its deviation from threshold levels (ideal, required, critical, etc.) or from baseline values associated with benchmark objects or situations exceeds a specified limit. Such trigger values may be defined on the basis of (i) aggregated direct observations of the indicator's components or (ii) forecast- and classification-based estimates.

A number of pressing challenges in indicator design and use arise when supporting decisions that concern the resource-based security of interests in the management of Organizational systems (OS) and in projects of national or international scale. Such decisions are tightly linked to limited — and often unstable — resource inflows, as well as to specialized resource categories whose availability depends on external interactions and innovation aimed at creating the necessary resource base [11].

The most demanding requirements for relevant and effective indicators of resource-based security of interests (RSIs) are set by OS whose Problem area (PA) exhibits the following properties:

- coexistence of distributed interest-security mechanisms with hierarchical operational control;
- planning under uncertainty (including deep uncertainty);
- simultaneous evaluation of the resource state with respect to both planned and crisis management capabilities and to norms of routine activities.

These complexities stem, on the one hand, from the OS structure — whose organizational elements retain their own internal management mechanisms — and, on the other hand, from a highly dynamic environment that demands flexible interaction in the face of new threats and opportunities affecting the OS's strategic interests.

A characteristic representative of this PA multidimensional management class, hereafter denoted MDM, is activity in the military domain of national security. Developing adequate RSIs for decision support in MDM faces further specifics:

- two distinct tiers of interest security — a global decision center and an internal hierarchy of implementation centers;
- multi-faceted (goal-oriented) and multi-criteria (performance-oriented) interest systems held by stakeholders in the decision centers, which engage in diverse management interactions;
- high uncertainty of both current and future decision contexts, driven by interdependencies with other PAs, environmental volatility, and shifting internal priorities;
- the PA still lacks any *a priori* PA-specific model that formally relates interest-attainment KPIs to resource-availability indicators, even though empirical evidence leaves no doubt that this relationship is substantial.

Building on the above requirements and current trends in indicator design, we posit that effective resource-based security indicators (RSIs) must satisfy three design principles:

- **Compositeness** – each RSI should integrate components that are relevant to the different interest systems operating within the OS;
- **Informational sufficiency** – the model must rely solely on data that can be extracted from the OS business-process information environment, ensuring practical deployability;

- **Adaptivity** – the model should detect triggers in the current OS state that call for method selection and parameter adjustment, and respond accordingly.

A promising paradigm for constructing an RSI model is to treat resource situations as a classification problem and apply machine-learning methodology [12]. This choice offers the following advantages:

- the model is learned from documented operational history rather than from *a priori* theoretical PA knowledge;
- classes are defined by the stakeholder-accepted gradations of interest well-being that are already employed in expert decision making;
- the same information space — routine resource monitoring, post-operation assessments, and scheduled expert audits — supports parallel evaluation of resource favourability from multiple decision-center viewpoints;
- the RSI remains adaptive through incremental retraining and robust through dynamic selection and tuning of methods according to the statistical properties of observed resource states.

The study therefore aims to create conceptual knowledge models for the MDM PA that enable RSI construction with the above properties and to embed them in a composite, multi-functional indicator system for PA state assessment. Within that system the machine-learning RSI is developed to provide *express diagnosis of resource risk to strategic interests.*

2. Conceptual Model of Domain S_i within the MDM Problem Area

The conceptual model is organized in two levels

$$\{L_k\}\ k=1,2 \tag{1}$$

where L_1 – management of interest security;

L_2 – management of internal activities that support those interests.

Each level is specified by a structural-and-functional model:

$$SFM_k = (DC_k, BC_k, MB_k, IB_k, RCM_k, ERCM_k) \tag{2}$$

where DC – decision centers (DCs);

BC – advisory and situational communications that support distributed decision making;

MB – hierarchical management links;

IB – regulatory information links;

RCM – types of reference concepts of management;

$ERCM$ – reference concepts of management.

Decision-center model

Every DC denoted $dc_{kl} \in DC_k$ (see (2)) is described by

$$Mdc_{kl}=(T, RT, A, SUB, \{(MI, idc, ircm)\}, hrcm) \tag{3}$$

where T – functional type;

RT – role type;

A – set of activity types;

$SUB \in ERCM$ – closest higher-level DC;

MI – management impact on another center idc by means of the impact concept $ircm \in ERCM$;

$hrcm \in ERCM$ – concept that represents OS interests in domain S_i for this DC.

The scheme supports the coexistence of hierarchical cascading [13], distributed [14] and deliberative [15] decision processes, and allows combining agile strategic planning [16] with capability-based planning [17].

Within each structural-and-functional model SFM_k in (2) the set $ERCM_k$ captures the reference concepts of management. An individual concept $rcm \in ERCM$:

- formalizes a specific aspect of OS stakeholder interests,
- serves as the basic object of activity for a decision center of the corresponding functional type T in model (3),
- provides the information schema used to monitor domain S_i, and
- acts, depending on the DC level, either as a planning target at L_1 (strategic tier) or as the operationalized output of strategy at L_2 (operational tier).

Thus, $ERCM_k$ links the decision-center model Mdc_{kl} in (3) to the higher-level structure of SFM_k in (2), supplying the conceptual "glue" that harmonizes hierarchical, distributed and deliberative decision processes.

RCM element types

1. Strategic goal

$$SG_i = (\{GO_{ik}\}_{k=1,\ldots,K},\ GC_i,\ MG_i) \tag{4}$$

where GO – target object;
K – number of objects;
GC – condition for desired state;
MG – progress metric.

2. Strategic capability

$$CAP_i = (MT_i,\ MC_i, OC_i, \{RES_{ik}, NR_{ik}\}_{k=1,K}), MCAP_i, SSG_i) \tag{5}$$

where MT – task whose execution capability is evaluated;
MC – execution conditions;
OC – set of key DCs to which the capability is delegated (see model (3));
RES_k, NR_k – resource type and norm (number of resource types K in (5) depends on the chosen resource model [17]);
$MCAP$ – metric that evaluates the current level at which the CAP is being realized by the task performers;
$SSG \subseteq \{SG_i\}$ – subset of strategic goals that can be achieved through this CAP.

3. Crisis management task

$$TA_i = (PS_i,\ TCAP_i,\ PT_i,\ TR_i,\ TF_i,\ ET_i) \tag{6}$$

where PS – problem situation;
$TCAP$ – capabilities to be employed;
$PT \in DC$ – performer DCs;
TR – operational resource plan;
TF – deadline;
ET – performance evaluations.

4. Activity ledger for an L_2 DC

$$DCAT_i = (DCT_i \{(CTA_{ik}, R_{ik},\ RS_{ik})\}_{k=1,\ldots,3}) \tag{7}$$

where DCT – set of delegated capabilities;
CTA – activity category: 1 - routine, 2 - capability support, 3 - crisis;
R – performance rating;
RS – resource support of the activity.

$$RS_{ikl} = \{TRS_{ikl},\ NRS_{ikl},\ MRS_{ikl}\}_{l=1,\ldots,L} \tag{8}$$

where TRS – resource type;
NRS – norm or plan;
MRS – current supply level;
L – number of resource types.

Below in Table 1 we specify, for every DC type and level, its functional role, activity portfolio, managerial impacts, and the reference concept of management (RCM) that represents the center's interests (strategic goal, capability, crisis-management task, etc.). This structure captures both hierarchical and distributed management schemes, records resource- and information flows among centers, and links the realization of the strategic interests of S_i to ongoing operational activity at level L_2.

This table enables precise mapping of resources, activities, and interests across both levels of the OS while supporting agile integration of hierarchical and distributed decision-making processes.

Table 1
Decision-center (DC) system for the MDM problem area

PA Level	DC identifier	Functional type of DC	Main activities	Managerial impacts: Addressee	Managerial impacts: Content	Reference concept of management: Type	Reference concept of management: Handling
L_1	dc_{11}	Oversight and coordination of interests and strategies	• OS-interest management in domain S_i • Cross-domain liaison • Monitoring external environment • Threat/vulnerability diagnostics • Strategy development & update	D_{21}	• Capability requirements • Threats, vulnerabilities, problem situations • New opportunities	Strategic goal	Formulation, update, state assessment
L_2	dc_{21}	Operational-integration center	• Cascading strategic tasks • Capability formation • Crisis-task planning • Assessing operational interests & performance • Strategic alignment • Resource allocation & ad-hoc needs • Building hierarchies $\{dc_{2lr}\}_r$ for operational needs	Key DCs $\{dc_{2l}\} \subset DC_2$ (see (2))	• Delegated capabilities • Crisis management tasks • Handling requests & resolvlicts • Subordination to key DCs within operational needs	Strategic goal Crisis management task	• iterative co-development with dc_{11} • assessment & correction Anticipatory planning
L_2	dc_{2l}	Key DC for direction l	• Detailed crisis-task planning • Routine & capability-support planning • Resource distribution within subordinate hierarchy • Evaluation of executed actions ication & disclosure of ad-hoc needs	Supporting DC set for key DC l $\{dc_{2lr}\}_r$	• Task cascading • Resource distribution • Activity evaluation	Activity ledger	• Requests to adjust resource norms • Maintaining supply data • Recording performance ratings
L_2	dc_{2lr}	Supporting DC for key DC l	• Planning participation in operational actions • Coordinating support across multiple dc_{2l} hierarchies • Managing own resources	No addressee	No addressee	Activity ledger	• Maintaining resource data • Recording and translicts and ad-hoc needs

3. Composite Multi-Functional Indicator System for the PA

To provide expert-analytic support for decision making in the Organizational system (OS) described in the previous section, we introduce an Indicator System for Interest Well-Being (*SI*) that follows three design principles drawn from modern composite-indicator research (see Sect. 1):

1. **Each constituent indicator** must quantify the well-being level of a specific reference concept of management (RCM), which serves as the constructive representation of an interest.
2. **Multiple decision-support functions** imply that constituent indicators must be compositely integrated into several resulting indicators, each oriented toward a distinct function.
3. **Inter-interest dependencies** — expressed through relevance relations $REL(tk_1, tk_2)$ among RCM types tk_1, tk_2 — govern whether an RCM of one type can influence, or supply data to, an RCM of another type.

Using the RCM models defined in (4)–(8) we formalize relation predicate $REL(O_r, O_s)$. These relations determine whether an RCM of one type can influence an RCM of another type or serve as an information source for integration. An RCM of a given type is referred to by the identifier introduced in (4)–(8), and an element A of the model of the s-th concept is denoted A_s.

$$REL(SG_r, CAP_s) \Leftrightarrow SG_r \in SSG_s \tag{9}$$

$$REL(TA_r, CAP_s) \Leftrightarrow CAP_s \in TCAP_r \tag{10}$$

$$REL(DCAT_r, TA_s) \Leftrightarrow DCAT_r \in PT_s \tag{11}$$

$$REL(SG_r, TA_s) \Leftrightarrow SG_r \in PS_s \tag{12}$$

$$REL(SG_r, DCAT_s) \Leftrightarrow (\exists\, TA_k : REL(DCAT_s, TA_k) \wedge REL(SG_r, TA_k)) \tag{13}$$

$$REL(CAP_r, DCAT_s) \Leftrightarrow CAP_r \in DC_s \tag{14}$$

Interest well-being in the performance aspect: direct estimate

The well-being of an RCM *oc* of type *TK* is represented by two components

$$MWB(TK,t) = (MWBR(TK,t), MWBS(TK,t)) \quad (15)$$

where *MWBR* – model of the performance aspect of the activities that sustain the interest represented by *oc*;

MWBS – model of the resource-support aspect of those activities.

Focusing on the first component,

$$MWBR(TK, t) = (\{A_i, ER_i, MEX_i, SC_i, IC_i\}_{i=1,\ldots 3,}, FIA) \quad (16)$$

where A_i – activity category (see (7));

ER_i – expert rating of performance;

MEX_i – expert-evaluation model used [18, 19, 20];

SC_i – rating scale;

IC_i – information context of the evaluation;

FIA – estimates integration function across activity categories.

Model (16) interprets the attained performance well-being as the degree to which stakeholder needs are satisfied, using expert ratings on a verbal–numeric scale [18] that matches their business practice. Thus

$$EWBR(oc,t) = FIA(ER_1, ER_2, ER_3) \quad (17)$$

is the direct estimate of the performance-aspect well-being of the interest represented by RCM *oc*.

Interest well-being in the performance aspect: indirect estimates

An alternative to the direct score (17) is a pair of indirect estimate sets

$$(\{EMRI\}, \{EMRS\}) \quad (18)$$

where *EMRI* – integration-based estimates;

EMRS – statistically derived estimates.

The integration estimate $EMRI_k(oc,t)$ is obtained by aggregating the direct scores *EWBR* of all RCMs of the *k*-th type that are relevant to *oc* but differ from its own type *TK*.

$$EMRI_k(oc,t) = FIMR_k(\{EWBR(rc_{kj},t)\}_{j=1\ldots J}) \quad (19)$$

where $\forall\ j \in (1,\ldots,J)\ REL(oc, rc_{kj})$;

FIMR – denotes the score-integration function.

Statistically derived estimates are built on a mass-observation array that records (i) the states of every RCM in the OS and (ii) the behavior of the external environment. The procedure constructs and applies machine-learning models that capture the relationship between performance well-being of a given target interest and the resource-support well-being that is observed or inferred for other RCMs — either formally relevant to the target or hypothetically exerting an influence on it.

We formalize this data as an observation array *MAOBS*(*TT*), which spans the entire observation period $TT = (TT^1, TT^2)$.

First, we introduce several special time sub-intervals inside the overall window *TT*:

- *T1* – the constant span between the *j*-th and (*j* + 1)-st scheduled audits of strategic-level RCMs (*SG* and *CAP*). The concrete instance of this span is denoted $(TT1_j^1, TT1_j^2) \subset TT$;
- $\{\{T2_{ij}\}_{i=1,\ldots N}\}_{j=1,\ldots M}$ – the set of durations of those crisis-management tasks ta_{ij}, that are carried out within the interval $(TT1_j^1, TT1_j^2)$;
- *T3* – the regulatory update period for each activity ledger *DCAT* its *j*-th occurrence is the interval $(TT3_j^1, TT3_j^2)$.

With these notations the observation array is expressed as

$$\begin{gathered} MAOBS(TT) = (\{EWBR_{jk}(SO_k, TT1_{kj}^1)\}_{k=1,K}\}_{j=1\ldots J1}, \{\{EWBR_{jl}(OTA_l, TT2_{lj}^1)\}_{l=1\ldots L}\}_{j=1\ldots J2}, \\ \{EWBR_{jm}(DCAT_m, TT3_{mj}^1)\}_{m=1\ldots M}\}_{j=1\ldots J3}, \{EWBS_{jk}(SO_k, TT3_{kj}^1)\}_{k=1\ldots K}\}_{j=1\ldots J3}, \\ \{\{EWBS_{jl}(OTA_l, TT2_{lj}^1)\}_{l=1\ldots L}\}_{j=1\ldots J2}, \\ \{EWBS_{jm}(DCAT_m, TT3_{mj}^1)\}_{m=1\ldots M}\}_{j=1\ldots J3}, \{\{R_{jr}\}_{r=1\ldots NF}\}_{j=1\ldots J1}) \end{gathered} \quad (20)$$

where *EWBR* – direct performance-aspect well-being score;

EWBS – resource-support well-being score;

SO_k – strategic-level RCM;

OTA_l – crisis-management task;

$DCAT_m$ – activity ledger of the *m*-th L_2 decision center;

$J1$ – number of audits;

$J2$ – number of crisis tasks executed;

$J3$ – number of *DCAT* content updates.

As an observation element for external-environment factors in the *MAOBS* model, we include the risk produced by the *r*-th external factor $\{F_r\}_{r=1...N}$ during the interval ($TT1_j^1$, $TT1_j^2$). This risk is evaluated by

$$R_{jr} = (L_{ri}/T_j)\cdot P_{rj} \tag{21}$$

where L_{ri} – duration of the action caused by F_r within the *i*-th audit interval;

P_{rj} – estimate of the factor's intensity in that interval.

The statistically derived estimate *EMRS* (see (18)) uses the elements of (20) as input data and is specified as

$$MEMRS(oc,t) = (MF, MT, \{MML_j\}_{j=1...J}, \{CR_r\}_{r=1...N}, FIE) \tag{22}$$

where MF – feature model;

MT – target-variable model;

$\{MML_j\}$ – set of machine-learning models employed;

CR – statistical criterion used to assess the quality of the results produced by the chosen *MML*;

FIE – function that integrates the individual model outputs.

$$MF = \{F_i, \{SO_{ij}\}_{j=1...Ji}, CO(t), \{SC_{ik}\}_{k=1...Ki}\}_{i=1...N} \tag{23}$$

where F_i – feature;

$SO_{ij} \in MAOBS$ (see (20)) – an observation-array element that can serve as a source for the feature value;

$CO(t)$ – a predicate that specifies how SO_{ij} values are selected in time, depending on the position of moment *t* relative to the baseline time-intervals defined in (20);

SC_{ik} – admissible scales for feature;

N – total number of features.

Target-variable model can be defined as

$$MT = (TV, \{SOT_i\}_{i=1...M}, \{SYN(t, F_i)\}_{i=1...N}, \{SCT_j\}_{j=1...J}) \tag{24}$$

where TV – target variable;

SOT_i – an element of the observation array that corresponds to the *EWBR* evaluations of the RCM representing the target interest;

$SYN(t, F_i)$ – rule for time-synchronizing the target and feature observations;

SCT_j – the *j*-th candidate scale applicable to the values of the target variable.

Constructing an EMRS-class well-being estimate allows the system to handle the following sources of uncertainty:

- the evaluation moment *t* lies inside an audit interval that has not yet been completed;
- one or more scheduled audits were skipped in several preceding planning intervals;
- the direct score $EWBR(oc,t)$ conflicts with the integration-based score $EMRI_k(oc,t)$ determined for $t=TT1_j^1$ or $t=TT1_j^2$.

Interest well-being in the resource-support aspect: evaluation

The resource-support aspect in (15) treats well-being as the closeness of the actually available amount of a given resource at time *t* to its normative (planned) level.

Direct evaluation is made only for Activity-ledger RCMs, ($TK=DCAT$).

In this case (see (8))

$$EWBS(oc,t)=FIA(ES_1,ES_2,ES_3) \tag{25}$$

where FIA – is the integration function for the category-specific scores *ES*.

$$ES_i=FIS(\{(NRS_{ij}-CRS_{ij})/NRS_{ij}\}_{j=1...J}) \tag{26}$$

where NRS_{ij}, CRS_{ij} – are, respectively, the normative and current resource supply;

FIS – integrates the adequacy levels across the J resource types.

For any $TK\neq DCAT$ the resource well-being of *oc* is obtained by integrating the direct scores of all relevant ledgers:

$$EMWBS_i(oc,t)=FIMS(\{EWBS(or_k)\}_{k=1...K}) \tag{27}$$

where each or_k – is a *DCAT* type RCM such that $REL(oc,or_k)$ holds (cf.(9)-(14));

$FIMS$ – the integration function over the *K* corresponding activity-ledger scores.

Structure and composition of the indicator system

Building a composite, multifunctional Indicator System for Interest Well-Being (*SI*) for an Organizational System (OS) operating in an MDM problem area rests on the constructive combination of:

- **individual indicators** *IC*, each describing the state of a single interest (these are the constituent indicators of the composite model discussed in Sect. 1);
- several **composite functional indicators** *ICF* that characterize the entire interest system and are defined as specific compositions of the individual indicators.

An individual indicator $IC_i(SO_j,t)\}$ represents the state of an RSM SO_j according to evaluation type i — direct (eqs. 17, 25) or integration-based (eqs. 19, 27).

Hence the overall indicator system is

$$SI(t)=(\{IC_i(SO_j,t)\}_{j=1,\ldots,N}\}_{i=1,\ldots,M},\ \{ICF_k\}_{k=1,\ldots4}) \tag{28}$$

where IC_i – individual indicator for RCM SO_j under evaluation type i;

ICF_k – one of the four basic composite functional indicators.

The core decision-support functions relevant to the PA activity model described in Sect. 2 comprise the following:

1. Rapid diagnosis of problematic situations;
2. Compilation of a vulnerability ranking;
3. Identification of threatening external factors and their potential targets within the PA;
4. Assessment of current strategy elements for consistency with up-to-date needs.

For direction k the functional indicator IF_k may rely on individual indicators and additional data. Formally, its model can be represented as

$$MICF_k(t)=(BIC_k(t),\ AD_k,\ PROC_k(BIC_k,AD_k),R_k) \tag{29}$$

where $BIC_k(t) \subseteq \{\{IC_i(SO_j,t)\}_{j=1,\ldots,N}\}_{i=1,\ldots,M}$ — the subset of individual indicators used;

AD_k – additional information, including formalized knowledge about relationships between PA elements or observations not captured by any *IC* but present in dataset (20);

$PROC_k$ – the procedure that calculates the indicator values;

R_k – the output vector (with defined scales) intended for use in decision making.

Table 2
Composite functional indicators of interest well-being

Functional indicator	Purpose	Constituent indicators	Composition: Additional information	Composition: Procedure	Output
1) Express diagnosis of resource risk to strategic interests	Rapidly detect potential problem situations for OS's strategic interests caused by resource shortfalls	• Direct & integrated performance scores for the target interest in previous periods • Direct resource-support scores for relevant DCs in the synchronous period	Results of statistical hypothesis tests on the stationarity of resource impacts	iers; apply them to obtain a statistically grounded indirect performance score	Binary well-being estimates plus quality metrics
2) Risk-exposure ranking	Identify resource types and institutional interest-holders exposed to resource risk	• Direct & integration-based resource-support scores by resource type & holder • Direct performance scores by activity category	Network of resource inter-relations between OS units	Multivariate time-ication of performance well-being versus resource well-being (model building and training)	Matrix of verbal–numeric well-being scores mapped to holders & resource types
3) Criticality of external factors	Reveal external factors most threatening to crisis tasks and the units most vulnerable to them	• Direct resource-support scores by activity category • Direct performance scores for crisis-management tasks of key DCs • Indirect integrated performance scores for strategic capabilities	History of external-factor manifestations from observation array (20)	Building the multi-iers of performance well-being vs. resource well-being & factors; sensitivity analysis	• Forecast ranking of fac-luence for the period, condi-ied resource well-being scores • Vulnerability ranking of operational-level elements and capability types

4) Strategic-adaptation diagnostic	Detect strategy elements that may require strategic alignment	All types of constituent indicators for all RCMs	Strategic-consistency model with: A) Normative rule set for well-being relations between RCMs & relevance links; B) hypotheses concerning non-compliant strategy elements and schemes for their plausibility assessment, derived from violations of rule set (A)	Determination of icance level of rule violations observed at the time of audit; derivation of the current hypothesis set and calculation of certainty factors [20].	Hypotheses on strategic-adaptation targets with plausibility levels

In Table 2, four functional indicators are described in accordance with structure (29).

The functional indicator *Express diagnosis of resource risk to strategic interests* — detailed in the next section — returns as its output an indirect performance well-being estimate of the given RCM *so* at time *t*, computed in the *EMRS* class (see (22)).

The elements of its model (22) are specified below.

Feature model *MF* (cf. (23)) comprises:

- features F_i, each is, as an element of *MAOBS* (20), the direct resource-support well-being score *EWBS*(*rso,to*) drawn from the DC's Activity ledger of the operational-level *rso* and fixed for a particular activity category and resource type;
- predicate *CO*(*t*) selecting *EWBS*(*rso,to*) values under the condition

$$REL(so,rso) \wedge (to \in (TT^1,t)) \tag{30}$$

Within *MF* the candidate scales *SC* include both (i) the standard continuous scale for *EWBS* on (0,1) and (ii) a scale of absolute resource quantities, the latter being used when the adopted normative values are deemed unreliable.

The target-variable model *MT* (24) permits only a binary scale, enabling indicator construction even when expert assessments are produced under deep uncertainty [20].

The temporal synchronization condition *SYN* for aligning target values *TV* with feature observations F_i in (24) is

$$\forall(to_k \in (TT1_j^1,\ TT1_j^2))\ TV(so,to_k)=TV(so,TT1_j^2) \tag{31}$$

Thus, every target-variable observation whose feature values fall within the window ending at moment ⊠ is assigned the direct performance well-being score produced by experts at the close of the corresponding well-being audit of *so*.

The procedure *PROC* specified in (29) for this indicator is presented in Section 4.

4. Construction of the *Express Diagnosis of Resource Risk to Strategic Interests* Indicator

This section presents the algorithm used to build and to forecast the resource-based security state of strategic goals. The task is solved with binary-classification models. The workflow comprises two major phases: model building (training) and forecasting.

4.1 Model-building (training) phase

Step 1. Preparation of the training set

A. Period selection and data synchronization

The training set is formed according to the feature model (23). Each input feature F_i is a numerical score of resource support *EWBS* for the corresponding Decision Center (DC) over the interval between two scheduled audits of the strategic goal. The predicate *CO(t)* aligns the time-stamp of *EWBS* with the moment of the expert performance well-being score *ESG*(T_a), which serves as the binary target variable (1 = satisfactory, 0 = unsatisfactory).

Every feature F_i can be represented in several alternative scales SC_i. For resource indicators we employ at least two:

- normalized *EWBS* ∈ (0,1);
- absolute resource quantity.

When required, the normalized *EWBS* may be discretized into categorical levels; this eases expert interpretation and broadens the family of machine-learning models that can be applied. The absolute scale is preferable when the planned norms are considered potentially outdated, so that comparing against the raw values is more reliable. Maintaining several scales SC_i increases model flexibility and ensures proper comparison of resource metrics under heterogeneous conditions.

B. Exploratory data analysis

a) Sample size and completeness

The dataset is examined for sufficiency with respect to the requirements of the intended model families. Missing values are handled either by deletion or by an appropriate imputation strategy, chosen after assessing both feasibility and impact on data integrity.

b) Distributional form and feature independence

For every variable we test compliance with theoretical distributions assumed later in the pipeline (e.g., normal, log-normal) and identify pairs of highly correlated features. Whenever the detected correlation exceeds an admissible threshold, the offending features are either decorrelated (by transformation) or removed.

c) Balance of the target variable

The ratio of *satisfactory* to *unsatisfactory* cases for $ESG(T_a)$ is computed. If a pronounced skew is observed, the training set is rebalanced through oversampling/undersampling or by using class-weighted loss functions to prevent majority-class bias. Because most machine-learning algorithms are sensitive to imbalance, subsequent classification-quality metrics are calculated in a way so that unequal class distributions are properly reflected.

Step 2. Training the full set of candidate classifiers and generating base classifications

Because no single learning algorithm is universally optimal (the No-Free-Lunch theorem (NFL) [21]) the indicator is built on a heterogeneous ensemble of learners. Candidate models were selected from the families surveyed in [22] according to four criteria:

- representation of the main classification paradigms (statistical, geometric and logical);
- compatibility between the algorithm's data-volume requirements and the size of the available resource dataset;
- documented evidence of successful practical use;
- availability of efficient software implementations.

The resulting set $PC = \{ MC_i \}$ comprises five classifiers: Bayesian classifier, Support-Vector Machines (SVM), Random Forests, k-Nearest-Neighbours (kNN) and Logistic Regression.

For every classifier $MC_i \in PC$ an expert characterization of the criticality of four data requirements was compiled from an extensive literature review [12, 23–53]:

$$KD_i = \{kd_{ij}\}_{j=1,\ldots,4} \tag{32}$$

where $j = 1$ – feature-distribution;

$j = 2$ – feature independence;

$j = 3$ – observation data balance;

$j = 4$ – measurement scale.

Each requirement is graded on a three-level expert scale $ekd_{ij} \in \{0, 0.5, 1\}$, where 0 – no requirement, 0.5 – violations admissible, 1 – requirement critical. Table 3 summarizes the resulting scores and provides the supporting rationale. For every criterion *j*, four argumentation aspects A_{aj} are documented (with $a = 1,2,3,4$): 1 – existence of a primary limitation; 2 – available robustness techniques; 3 – availability of software implementing these techniques; 4 – potential risks associated with their application.

Table 3
Critical-requirement scores ekd_{ij} for data-requirement classes with supporting rationale

	Feature-distribution A_{a1}	Feature independence A_{a2}	Observation data balance A_{a3}	Measurement scale A_{a4}
		ier		
ekd_{i1}	**0**	**0**	**0.5**	**0**
$a = 1$	ication of the underlying probability model is required	Assumption of conditional independence between features	ication quality	Model must adapt to the data-type scale
$a = 2$	Dirichlet-process mixture models can be used to let the data dictate distributional form	Dependencies can be captured via multivariate or conditional distributions	Priors, regularization, and scaling to mitigate imbalance	Choose appropriate proba- ferent feature types
$a = 3$	R packages *brms*, *DirichletProcess*	*brms* (R) supports dependent features	R packages *brms, caret*	*brms* (R) supports diverse data types
$a = 4$	Minimal	None; modelling dependencies increases accuracy	Possible information (explanatory power) loss after scaling	None
		SVM		
ekd_{i2}	**0**	**0**	**0.5**	**0.5**
$a = 1$	None; classical linear SVMs fectively separate classes with complex distributions	None; optimization might be required under strong feature correlation	Inherent bias toward the majority class	Highly sensitive to feature scale
$a = 2$	Kernel functions allow modelling of complex non-linear distributions	Mapping data into higher-dimensional feature spaces via kernels	Class weights, resampling methods, SMOTE	Data scaling (normalization, standardization)
$a = 3$	R packages *e1071*, *kernlab*, *caret* provide SVM with various kernels	R packages *e1071*, *kernlab*, *caret*	R packages *DMwR*, *ROSE*, *caret*	The *caret* package supplies extensive preprocessing utilities
$a = 4$	Minimal; method is generally robust to distributional assumptions	Minor; impact is limited to computational cost	Potential loss of discriminative power	Non-critical constraints
		Random Forest		
ekd_{i3}	**0**	**0**	**0.5**	**0**
$a = 1$	None	None	Bias toward the majority class	None
$a = 2$	Not required	Random feature sampling at splits mitigates correlation impact	Class weights, resampling	Not required
$a = 3$	R packages *randomForest*, *ranger*, *randomForestSRC*, *imbalanced*			
$a = 4$	None	None	Possible loss of discriminative power	None
		k-Nearest Neighbours (kNN)		
ekd_{i4}	**0**	**0.5**	**0.5**	**0.5**
$a = 1$	None	Feature dependencies distort distance-based results	Bias toward the majority class	Distance metric is scale-sensitive
$a = 2$	Not required	Dimensionality reduction (PCA, t-SNE)	Class weights, resampling techniques	Normalization, standardization
$a = 3$	R packages *caret*, *ROSE*, *DMwR*, *kknn*			
$a = 4$	Minimal	Potential loss of discriminative power	Loss of discriminative power	Wrong scaling shifts results
		Logistic Regression		
ekd_{i5}	**0**	**0.5**	**0.5**	**0.5**
$a = 1$	None; non-normality can weaken reliability	Multicollinearity among predictors	Bias toward the majority class	Requires correct encoding of categorical variables
$a = 2$	Transformations (logs, roots) to handle outliers	Remove correlated features, PCA, L1/L2 regularization	Class weights, resampling	fect coding, ordinal encoding
$a = 3$	Base R *stats::glm()*; package *caret*			
$a = 4$	Minimal	Possible loss of discriminative power	Possible loss of discriminative power	Incorrect encoding may reduce accuracy

The argumentative ratings presented in Table 3 enable subsequent steps of the algorithm to deploy each classifier within an adaptive ensemble model that reflects the problem area's specific characteristics across the four assessed dimensions. These ratings support (i) the use of the indicator ensemble when the statistical properties of incoming data are still uncertain, and (ii) timely detection of the need for retraining — triggered when the current weight coefficients assigned to the ensemble's classifiers deviate significantly from those learned earlier.

Step 3. Computing the trust degree of classifier

For every classifier *i* a model-data conformity coefficient KDR_i is calculated

$$KDR_i = \frac{1}{4}\sum_{j=1}^{4} ekd_{ij} \cdot IV_{ij} \quad (33)$$

where $IV_{ij} \in \{0,1\}$ — Incidence Variable denoting whether requirement *j* is violated (1) or not (0) for the current data set.

A larger KDR_i signals more severe violations of the classifier's underlying assumptions and therefore lower suitability. Conversely, the value $(1- KDR_i)$ is interpreted as the *trust degree* of classifier *i* for the given data.

Step 4. Integral classification-quality assessment

For every classifier an aggregated quality score KQ_i is computed. It combines several metrics that are robust to class-imbalance — Cohen's Kappa, F1-score and MCC. A higher KQ_i indicates that the model reproduces the true state labels more accurately and consistently, even when the class distribution is highly skewed.

Step 5. Weighted voting

For every classifier an aggregated quality score

$$CMR = \langle RC_i, KDR_i, KQ_i \rangle_{i=1}^{5} \quad (34)$$

where i — indexes the five base classifiers in the pool *PC*;

$RC_i \in \{0,1\}$ — is the class label returned by the *i*-th classifier at Step 2.

The *CMR* array is partitioned into two subsets according to the class predicted by each model

$$SP = \{CMR_i \mid RC_i = 1\},\ SN = \{CMR_i \mid RC_i = 0\},\ SP \cup SN = CMR \quad (35)$$

A weighted evidence index is then calculated for each subset:

$$IP = \frac{1}{|CMR|}\sum_{i=1}^{|SP|}(1 - KDR_i) \cdot KQ_i,\ CMR_i \in SP \quad (36)$$

$$IN = \frac{1}{|CMR|}\sum_{i=1}^{|SN|}(1 - KDR_i) \cdot KQ_i,\ CMR_i \in SN \quad (37)$$

where *IP* corresponds to the positive class (1) and *IN* to the negative class (0);

$|CMR|$ — is the total number of classifiers that cast a vote.

Step 6. Generating the final prediction

1. *Single-model (priority) mode*

The classifier with the highest combined weight is selected

$$MCP = \mathrm{MC_i} | max_{i \in 1,5}((1 - KDR_i) \cdot KQ_i) \quad (38)$$

and its binary output is taken as the final decision.

2. *Weighted-voting ensemble mode*

Using the evidence indices *IP* and *IN* defined in (36)–(37), the final class label is determined as

$$ICL = \begin{cases} 1, & IP > IN \\ 0, & IP < IN \\ undecided, & IP = IN \end{cases} \quad (39)$$

If $IP = IN$, the system returns *undecided* or, according to the operating protocol, requests an expert judgment.

4.2 Forecasting phase

After model training (Section 4.1) the forecasting phase repeats the same synchronization, verification, and feature-transformation procedures, then produces the class labels and maintains system adaptability.

The prediction phase supports two modes:

1. Single-model (priority) mode. A single classifier, selected by (38), is applied.
2. Weighted-voting ensemble mode. The outputs of the five classifiers are integrated according to (39).

This two-level design permits flexible integration of heterogeneous algorithms and delivers robustness to shifts in the distribution of input data. The resulting estimates can feed both automated monitoring services and strategic management decision support.

The forecasting pipeline is governed by the learning paradigm of each classifier. Parameter-based models — Bayesian classifier, SVM, logistic regression, and Random Forest — are applied directly with their stored parameters; retraining is required only if the exploratory data analysis (EDA) performed at inference time reveals substantial divergence from the data characteristics observed during initial training.

The non-parametric kNN classifier belongs to the lazy-learning family: for every incoming query it searches the full training set for the nearest neighbors and therefore needs no separate training stage [54, 55].

Algorithmic adaptability is maintained through continuous monitoring of the aggregate quality metric KQ_i. When the running average over the last N predictions drops by more than δ%, the system automatically triggers retraining of the parametric models and refreshes the instance base used by kNN. This mechanism preserves the stability and accuracy of the indicator system under shifting data distributions and emerging feature dependencies.

5. Conclusions

The composite, multi-purpose resource-based security indicator system (RSI) presented in this paper provides decision support for resource-driven choices in a macro-level organizational system that features multi-tier governance and a high uncertainty — both in the hostile external environment and in rapidly shifting internal priorities.

The RSI framework classifies resource situations with machine-learning methods trained on historical observations and expert judgements, while clearly representing stakeholder needs and environmental changes. For every strategic or operational interest, a set of alternative well-being estimates acts as an individual indicator; well-being is analyzed in two complementary aspects:

- Performance aspect — the degree to which the interest is achieved from the stakeholder perspective.
- Resource aspect — the degree to which available resources meet the normative and planned levels defined by the hybrid management schemes in use.

Four composite functional indicators are produced by integrating individual indicators and cross relating the well-being of different interests. They support resource-conditioned decision making in the following directions:

1. Express diagnosis of resource risk to strategic interests (rapid detection of problem situations);
2. Vulnerability ranking (prioritizing interests and resources at risk);
3. Criticality of external factors (assessing the impact of threats from the environment);
4. Diagnostic of strategic adaptation needs (identifying elements of the current strategy that may require realignment).

The algorithm for the first indicator — express diagnosis of problem situations — forecasts the performance well-being of strategic goals on the basis of routine resource monitoring, interim assessments of crisis-management tasks, and capability status. A heterogeneous ensemble of binary classifiers is employed; their outputs are merged through weighted voting, where the weight equals the argumentation index that combines (i) the data-conformity coefficient — how well the classifier's statistical assumptions match the observed data — and (ii) an aggregate quality score compiled from imbalance-robust metrics. This mechanism yields a robust indicator under prior uncertainty about data distributions and automatically triggers model retraining when conformity degrades.

Machine learning therefore becomes not merely a technical means of constructing the RSI but a core component of an intelligent decision-support platform for problem areas (PA) of the MDM class. Future work will extend the approach to the other composite indicators, employing multi-class classification and hybrid time-series forecasting, and will equip the indicator algorithm with a dedicated multi-criteria evaluation module that combines several imbalance-robust metrics into the aggregated quality score KQ_i. A concrete application is foreseen in the analytic integration of defense-resource information for decision support within the national-security domain.

Declaration on Generative AI

During the preparation of this work, the authors used OpenAI ChatGPT to translate text from Ukrainian into English and to rephrase sentences for improved clarity, conciseness, and style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, 2008. ISBN 978-92-64-04345-9.

[2] U. C. Schroeder, Measuring Security-Sector Governance – A Guide to Relevant Indicators. DCAF Occasional Paper 20, 2010, 61 pp.

[3] UNAIDS. Monitoring and Evaluation Fundamentals: An Introduction to Indicators. UNAIDS, 2010, 98 pp.

[4] D. Kaufmann, A. Kraay, The Worldwide Governance Indicators: Methodology and 2024 Update. World Bank Policy Research Working Paper, World Bank Group, Washington D.C., 2024. URL: http://documents.worldbank.org/curated/en/099005210162424110.

[5] C. Cavicchia, P. Sarnacchiaro, M. Vichi, G. Zaccaria, An ultrametric model to build a composite-indicators system. Book of Short Papers, IES 2022, pp. 208–211.

[6] O. I. Cherniak, M. I. Shumayeva, Methodology for constructing composite indicators. Visnyk of Zaporizhzhia National University. Economic Sciences 1 (21) (2014), 81–91 (in Ukrainian).

[7] Yu. M. Kharazishvili, Methodological approaches to assessing a country's economic security level. Science and Science of Science 4 (2014), 44–57 (in Ukrainian).

[8] A. B. Kachynskyi, Indicators of National Security: Definition and Application of Threshold Values. NISS, Kyiv, 2013 (in Ukrainian).

[9] E. Diener, E. M. Suh, Measuring quality of life: Economic, social and subjective indicators. Social Indicators Research 40 (1) (1997), 189–216.

[10] R. Klimaitienė, E. Derengovska, K. Rudzionienė, Application of key performance indicators to improve the efficiency of monitoring of the organisation's activities: theoretical approach. Public Security and Public Order 25 (2020), 218–233. DOI:10.13165/PSPO-20-25-20.

[11] I. P. Sinitsyn, V. L. Shevchenko, A. Yu. Doroshenko, R. M. Fedorenko: *Models and Software Systems for Management of Defense Resources*. Institute of Software Systems, NAS of Ukraine, Kyiv, 2024 (in Ukrainian), 268 pp.

[12] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009. ISBN 978-0-387-84857-0.

[13] M. Safari, M. Z. Mazdeh, A conceptual framework of strategy cascading in mission-based organisations: state-of-the-art review and practical template. International Letters of Social and Humanistic Sciences 83 (2018), 1–10. DOI:10.18052/www.scipress.com/ILSHS.83.1.

[14] C. Wernz, A. Deshmukh, Multiscale decision-making: bridging organisational scales in systems with distributed decision-makers. European Journal of Operational Research 202 (3) (2010), 828–840. DOI:10.1016/j.ejor.2009.06.022.

[15] O. Renn, Analytic–deliberative processes of decision making: linking expertise, stakeholder experience and public values. Research report Doc 847, 2004.

[16] L. Agostini, A. Nosella, R. M. Sarala, C. Nikeng, Emerging trends around strategic flexibility: a systematic review supported by bibliometric techniques. Management Decision 62 (13) (2023). DOI:10.1108/MD-02-2023-0135.

[17] J. Correia, Military capabilities and the strategic-planning conundrum. Security and Defence Quarterly 24 (2) (2019), 21–48. DOI:10.35467/sdq/108667.

[18] V. H. Totsenko, Methods and Systems for Decision Support: Algorithmic Aspect. Naukova Dumka, Kyiv, 2002, 381 pp. (in Russian).

[19] O. P. Ilina, Functions and methods supporting modern paradigms of the Delphi method. Problemy Programmirovaniya 1 (2009), 36–52 (in Russian).

[20] O. Ilina, I. Sinitsyn, O. Slabospitska, Models, methods and technological usage of expert-knowledge formalisation for strategic decision making under deep uncertainty. Proceedings of the 13th International Scientific and Practical Conference 'Programming UkrPROG' 2022, CEUR-WS Vol-3501, 2022, pp. 302–314.
[21] D. H. Wolpert, W. G. Macready, No Free Lunch Theorems for Optimization. IEEE Transactions on Evolutionary Computation 1 (1) (1997), 67–82. DOI: 10.1109/4235.585893.
[22] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real-world classification problems? Journal of Machine Learning Research 15 (2014), 3133–3181.
[23] J. M. Bernardo, A. F. M. Smith, Bayesian Theory. Wiley, 1994.
[24] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis. 3rd ed. CRC Press, 2013.
[25] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, New York, 2006.
[26] B. Schölkopf, A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2002.
[27] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, 2000.
[28] Y. Lin, Y. Lee, G. Wahba, Support vector machines for classification in nonstandard situations. Machine Learning 46 (1 – 3) (2002), 191–202.
[29] L. Breiman, Random Forests. Machine Learning 45 (1) (2001), 5–32.
[30] C. Chen, L. Breiman, Using Random Forest to Learn Imbalanced Data. Univ. of California, Berkeley, Tech. Rep., 2004.
[31] C. C. Aggarwal, Data Mining: The Textbook. Springer, 2015.
[32] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R. Springer, 2013.
[33] D. W. Hosmer, S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression. 3rd ed. Wiley, 2013.
[34] Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black, Multivariate Data Analysis. Prentice Hall, 1998.
[35] B. C. Wallace, K. Small, C. E. Brodley, T. A. Trikalinos, Class imbalance, redux. In: Proc. IEEE ICDM 2011, pp. 754–763. DOI: 10.1109/ICDM.2011.33.
[36] A. Gelman, J. Hill, A. Vehtari, Regression and Other Stories. Cambridge University Press, 2020.
[37] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques. MIT Press, 2009.
[38] C. M. Bishop, Neural Networks for Pattern Recognition. Oxford University Press, 1995.
[39] M. Buda, A. Maki, M. A. Mazurowski, A systematic study of the class-imbalance problem in convolutional neural networks. Neural Networks 106 (2018), 249–259.
[40] G. Wu, E. Y. Chang, KBA: Kernel boundary alignment considering imbalanced data distribution. IEEE TKDE 17 (6) (2005), 786–795.
[41] V. N. Vapnik, The Nature of Statistical Learning Theory. Springer, 1995.
[42] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines. ACM TIST 2 (3) (2011), Article 27.
[43] C.-W. Hsu, C.-J. Lin, A comparison of methods for multi-class support vector machines. IEEE Trans. Neural Networks 13 (2) (2002), 415–425.
[44] P. Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012, chap. 5.
[45] L. Rokach, O. Maimon, Top-down induction of decision-tree classifiers — A survey. IEEE Trans. SMC-C 35 (4) (2005), 476–487.
[46] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random-forest variable-importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8 (2007), Article 25.
[47] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, A. R. Feinstein, A simulation study of the number of events per variable in logistic-regression analysis. Journal of Clinical Epidemiology 49 (12) (1996), 1373–1379.

[48] R. Akbani, S. Kwek, N. Japkowicz, Applying support vector machines to imbalanced datasets. In: Proc. ECML 2004, LNCS 3201, Springer, 2004, pp. 39–50.
[49] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik, Feature selection for SVMs. Advances in Neural Information Processing Systems 13 (2001), 668–674.
[50] C. Cortes, V. Vapnik, Support-vector networks. Machine Learning 20 (3) (1995), 273–297.
[51] H. He, E. A. Garcia, Learning from imbalanced data. IEEE TKDE 21 (9) (2009), 1263–1284.
[52] H. He, Y. Ma (eds.). Imbalanced Learning: Foundations, Algorithms, and Applications. Wiley, 2013. ISBN 978-1-118-07462-6.
[53] J. M. Johnson, T. M. Khoshgoftaar, Survey on deep learning with class imbalance. Journal of Big Data 6 (2019), Article 27. DOI: 10.1186/s40537-019-0192-5.
[54] D. Wettschereck, D. W. Aha, T. Mohri, A review and empirical evaluation of feature-weighting methods for a class of lazy-learning algorithms. Artificial Intelligence Review 11 (1 – 5) (1997), 273–314.
[55] T. M. Cover, P. Hart, Nearest-neighbor pattern classification. IEEE Transactions on Information Theory 13 (1) (1967), 21–27.